Lipoprotein ontology as a functional knowledge base
Meifania Chen, Maja Hadzic
Research Lab for Digital Health Ecosystems, Curtin University of Technology, Perth, Australia
m.chen@curtin.edu.au, m.hadzic@curtin.edu.au

Abstract- The advances of high throughput research in the biomedical domain have resulted in an onslaught of data being generated at an exponential rate. As a result, researchers face challenges in navigating through overwhelming amounts of information in order to derive relevant scientific insights. Ontologies address these issues by providing explicit description of biomedical entities and a platform for the integration of data, thereby enabling a more efficient retrieval of information. There have been major efforts in the development of biomedical ontologies in the recent years; however no such ontology exists for lipoproteins, which play a crucial role in various biological and cellular functions. Dysregulation in lipoprotein metabolism is significantly associated with an increased risk to cardiovascular disease, the leading cause of mortality in the world today. The aim of this paper is to propose a preliminary framework for Lipoprotein Ontology, with particular focus on the etiology and treatment of lipoprotein dysregulation. This may provide a novel and effective strategy for managing at risk individuals.

I. INTRODUCTION

Lipoproteins are a water-soluble “lipid+protein” complex which serves as a mode of transport for fats and cholesterol from the intestines to the liver, as well as endogenously synthesised lipids from the liver to tissues. They consist of a hydrophobic core of triglycerides and cholesteryl esters and are surrounded by a hydrophilic outer layer of phospholipids, cholesterol and apolipoproteins. Lipoproteins play a very crucial role in the regulation of biological and cellular functions in humans, and can be impacted by a number of factors, including obesity, diet/nutrition, physical exercise and other factors such as smoking and alcohol consumption.

The complexities associated with lipoprotein dysregulation are continuously being revealed and addressed. However, as new information are constantly added to the massive pool of literature at a staggering rate, extracting relevant information proves to be a much harder challenge. Furthermore, lipoprotein research is an extensive area with a myriad of different research groups focusing on various aspects.

Ontologies serve as a semantic framework which can be used for systematic annotation of information which can be derived from heterogeneous sources. This annotation allows for better understanding of the content of the information source as well as the intelligent retrieval of relevant information. Ontologies can also be employed as robust knowledge bases which will facilitate collaboration between different research groups.

During the recent years there have been major efforts in the in the organising of biological concepts in the form of controlled terminologies or ontologies [15, 16, 17, 18]. The Unified Medical Language System (UMLS) is a repository of biomedical vocabularies which facilitates information sharing among various terminology systems [19]. The UMLS consists of three knowledge sources: Metathesaurus, Semantic Network, SPECIALIST Lexicon, as well as a number of supporting software tools. There are over 1 million biomedical concepts in Metathesaurus classified by 135 semantic types and 54 relationships. Additionally, various biomedical ontologies in various disciplines are available via The Open Biomedical Ontologies in a standard format that allows systematic updating on the basis of community feedback [20]. The Gene Ontology (GO) project provides a set of dynamic, controlled vocabularies of gene products that can be applied in different databases to annotate major repositories for plant, animal and microbial genomes [17]. Protein Ontology annotates terms and relationships within the protein domain and classifies that knowledge to allow reasoning [21]. Most recently, Lipid Ontology was developed to provide a structured framework for the effective derivation of lipid-related information [22]. Lipid Ontology mainly serves as a formal annotation for the classification and organisation of information on lipids, with minor inclusion of the implications of disease on lipids.

To date, there does not exist a formal ontology for lipoproteins. By creating annotations linking primary data to expressions in controlled, structured vocabularies, Lipoprotein Ontology will have a positive impact for ongoing lipoprotein research by making the data available to effective searching and algorithmic processing. Lipoprotein Ontology provides a formal framework for lipoprotein concepts and relationships that can be used to annotate database entries and support an intelligent retrieval of information. In addition, Lipoprotein Ontology enables efficient sharing of information, thereby promoting collaboration among various lipoprotein research groups. A key difference between Lipoprotein Ontology and other related ontologies is that Lipoprotein Ontology is developed with practical implications in mind. Lipoprotein dysregulation is becoming an increasingly serious concern with dire implications on health. By integrating all aspects of lipoprotein research, we can harness that knowledge into problem solving tools. As a result, Lipoprotein Ontology will not only provide a more efficient way of retrieving relevant information for the intended users, it will also support the querying of heterogeneous information sources from different aspects of lipoprotein research to achieve a common aim.

II. SIGNIFICANCE OF LIPOPROTEIN ONTOLOGY WITHIN THE BIOMEDICAL DOMAIN

Experienced scientists, clinicians and professionals are overwhelmed with huge, rapidly accumulating amounts of information on lipoproteins. There needs to be a systematic
approach to making use of an enormous amount of available information that has no value unless analysed and linked with other available information from the same domain. The three main problems are [23]:

1. The amount of the existing information. The size of the existing corpus of knowledge on lipoproteins is very large. The possibility of searching this information effectively is very low and in most cases some important information are neglected.

2. The dispersed nature of the information. Information regarding lipoproteins is dispersed over various resources and it is difficult to link this information, to share it and find specific information when needed. Additionally, the information contains a huge range of results.

3. The autonomous nature of the information sources. Most of the information resources develop their content independently. As a result of this, information available through different resources is heterogeneous in its format and content. This makes collective analyses of data very difficult. Additionally, many overlaps and redundancies are found in the data originating from different sources.

Ontologies are increasingly being used within the biomedical domain to bring some order into this chaotic situation of exponential information growth. A large number of biomedical ontologies has already been developed, and these ontologies have been used in various ways. We have identified five key application areas for ontologies:

1. Ontology-based collaboration between research teams
2. Ontology-mediated information access, management and retrieval
3. Ontology-based multi-agent systems
4. Ontology-aided data mining
5. Ontology-based web services

Ontologies are useful in situations where collaboration between different research teams needs to be established in order to share complementary information and build a joint solution to their shared problem. Using an ontology to formalise the common knowledge will also reduce the possibility of duplicating experiments by different research groups (such as examining the same region of a DNA sequence), thus saving time and resources. This will create a cooperative environment, allowing coherence between different research teams in big research tasks. Gap analysis in lipoprotein research will also be easier. Research projects would be mapped onto the lipoprotein domain knowledge as specified by the ontology. During the planning stage, researchers could use these maps to see where overlaps, redundancies, complementary research and (more importantly) no research has occurred or is occurring.

Lipoprotein Ontology can be designed to provide a conceptual model of lipoprotein concepts and relationships. This model can be used to form a semantic framework for information access, management and retrieval tasks. Lipoprotein Ontology can be used for systematic annotation of lipoprotein information available through various information resources. The annotation of information enables searching for the meaning of information over its appearance in the text, and in this way, allows for intelligent querying of heterogeneous information sources and retrieve relevant information. Most of the current search engines do not operate in this way as they search for a particular string of the letters and not for the meaning of the word. Consequently, such search engines often retrieve voluminous results that are completely outside the domain the user is interested in. Furthermore, ontologies can also be used as mediators to bridge the heterogeneities between different data resources, information services and user applications.

Powerful systems can be designed when ontologies are integrated with multi-agent technologies [24]. Ontologies can be used as coherent and consistent knowledge bases of different agents. A collection of agents can utilise shared domain ontology as their common knowledge base. This will facilitate communication and coordination between agents and support some important processes within a multi-agent system such as: problem decomposition and task sharing among different agents, results sharing and analysis, information retrieval, selection and integration. Ontology and multi-agent technologies can be used together to enable efficient and effective access, extraction and manipulation of the information from various information resources.

Data mining is a set of processes that is based on automated searching for actionable knowledge buried within a huge body of data. Data mining extracts information and finds patterns and behaviours embedded in the data. Lipoprotein Ontology is machine-understandable and allows automatic data analysis programs such as data mining to access, understand and analyse the ontology-structured information. When applied on lipoprotein data, data mining algorithms can play a crucial role in deriving new knowledge within the lipoprotein domain. The results have the potential to increase the understanding of the dynamics involved in the lipoprotein pathways and help making predictive models for decision making and new discoveries. This new knowledge can be used to assist in the (1) prevention, diagnosis, treatment and control of dyslipidemia, (2) improvement of the infrastructure for evidence-based interventions and (3) introduction of innovations for improvement in health care.

The importance of ontologies within the Web Services community is significant as ontologies can provide a semantic framework that will enable knowledge sharing and use by all parties involved. In some of our works [25, 26] we have explained how Ontology technology in combination with Grid technology can help create collaborative environments and increase interoperability within the system.

III. LIPOPROTEIN ONTOLOGY MODEL

As lipoprotein research is an extensive area with a multitude of different research groups working disparately to achieve different aims, the need for a common information repository is warranted in order to fully appreciate the implications of lipoprotein dysregulation. By incorporating specific aspects of lipoprotein research in Lipoprotein
Ontology, not only in terms of the classification of lipoproteins, but also understanding the metabolic pathways, pathophysiology, causes and treatment of abnormal lipoprotein levels, this impacts not only on identifying the risks, but also provides effective preventative measures.

In this paper, we propose a preliminary model for Lipoprotein Ontology using Protégé, which consists of five sub-ontologies:

1. Classification – different classes of lipoproteins
2. Metabolism – lipoprotein metabolism pathways
3. Pathophysiology – understanding the interactions between various lipoproteins crucial to understanding lipoprotein dysregulation
4. Etiology – establishing the causes that might cause lipoprotein dysregulation
5. Treatment – management of lipid dysregulation

An overview of Lipoprotein Ontology is shown in Fig. 1. Each of the five sub-ontologies of Lipoprotein Ontology is described below.

A. Classification
Lipoproteins are classified according to size and density, i.e. chylomicrons, very-low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipoproteins (LDL), high density lipoproteins (HDL), ranging from largest and least dense to smallest and most dense [27].

Apolipoproteins are lipid-binding proteins which serve as enzyme modulators and receptor ligands that regulate the intravascular metabolism of lipoproteins and their tissue uptake.

B. Metabolism
Lipoprotein metabolism is divided into two pathways, exogenous, the transport of dietary lipids from the intestines to the liver and peripheral tissues, and endogeneous, the transport of lipids produced in the liver to peripheral tissues. As lipoprotein metabolism is highly complex, involving a number of different pathways, this will be further elucidated in our future work.

C. Pathophysiology
Dyslipidemia is defined by abnormal levels of lipoproteins. Some characteristics of lipoprotein dysregulation include high concentrations of low-density lipoproteins (LDL) cholesterol, low high-density lipoproteins (HDL) cholesterol and high plasma triglycerides [3].

D. Etiology
Possible causes of lipoprotein dysregulation are shown in Fig. 2.

Fig. 2. Ontology model of etiology of lipoprotein dysregulation. Sub-ontologies include genetics, lifestyles, which are further sub-divided into exercise, diet, smoking alcohol and emotional wellbeing, as well as drug interactions, which are further sub-divided into corticosteroids and beta-blockers.

Lifestyle contributes to most cases of dyslipidemia in adults. The most prominent cause in industrialised countries is a sedentary lifestyle with lack of physical exercise and excessive dietary intake of saturated fat and cholesterol. Other common causes include smoking, alcohol consumption and stress [28, 29].

Drug interactions may also contribute to lipoprotein dysregulation. Some examples are corticosteroids [30,31] and beta-blockers [32].

Genetic mutations may result in either the overproduction or defective clearance of LDL cholesterol and triglycerides, or the underproduction or excessive clearance of HDL [33].

E. Treatment
Treatment options for dyslipidemia are illustrated in Fig. 3.

Fig. 3. Ontology model of treatments for lipoprotein dysregulation. Sub-ontologies include combination therapy, drugs, further sub-divided into statins, fibrates, bile acid sequestrants and cholesterol absorption inhibitors, as well as lifestyle changes, further sub-divided into weight loss, dietary supplements and physical exercise.
First line approach includes lifestyle changes such as weight loss, exercise and dietary modification [2]. Dietary supplements include the following:

**Fish oil** has been reported to reduce plasma triglyceride concentrations by up to 40% [34]. Increasing evidence suggest that fish oil consumption protects against coronary disease [35, 36].

**Nicotinic acid (niacin)** in vitamin B₃ blocks the metabolism of fats in adipose tissue, leading to a decrease in free fatty acids in the blood and decreased hepatic secretion of VLDL and cholesterol [37]. Niacin also increases HDL by promoting HDL production and inhibiting HDL clearance.

**Antioxidants** inhibit the oxidation of toxic LDL and is inversely correlated with cardiovascular risk. Examples of antioxidants are vitamin A, vitamin E and vitamin C.

**Dietary supplements** that lower LDL cholesterol levels include fiber supplements and products containing plant sterols (phytosterols). Plant sterols reduce LDL cholesterol by 10%-15% by inhibiting cholesterol incorporation into micelles, decreasing the absorption of total cholesterol [38].

In addition to lifestyle changes, several lipid-regulating agents may be used to improve dyslipidemia [39, 40]. **Statins** are possibly the treatment of choice for reducing LDL cholesterol as they demonstrably reduce cardiovascular mortality. Statins inhibit hydroxymethylglutaryl CoA reductase, a key enzyme in cholesterol synthesis, leading to up-regulation of LDL receptors and increased LDL clearance [41, 42]. They reduce LDL cholesterol by up to 60%, produce small increases in HDL and modest decreases in triglycerides through the reduction of VLDL [12, 43]. Recent clinical trials have demonstrated that statins can decrease cardiovascular events, irrespective of the initial level of cholesterol [43, 44].

**Fibrates** reduce triglycerides by about 50% and significantly lowers VLDL content. Clinical studies have shown that fibrates can reduce cardiovascular events in high-risk subjects [45, 46]. The mechanisms of action of fibrates on lipoprotein metabolism have been elucidated in various experimental studies [47].

**Bile acid sequestrants** have been proven to reduce cardiovascular mortality. They prevent the reabsorption of intestinal bile acid, forcing the up-regulation of hepatic LDL receptors to recruit circulating cholesterol for bile synthesis, leading to a decrease in LDL levels [48].

**Cholesterol absorption inhibitors** inhibit intestinal uptake of both dietary and hepatic cholesterol. They lower LDL cholesterol by 15 to 20% and causes small increases in HDL and a mild decrease in triglycerides [49]. Cholesterol absorption inhibitors have been reported to have a potent effect when combined with statin therapy, and is extremely beneficial for statin-resistant or statin-intolerant populations.

**Combination therapy** involves the use of dual or multiple lipid-regulating agents to treat lipoprotein abnormalities by targeting specific lipoproteins and utilising the complementary mechanisms of action of the different agents [50]. This is a more effective treatment strategy where lipid-regulating monotherapy (e.g. statins or fibrates) may not provide adequate improvement in dyslipidemia. However, there are some contraindications between different treatments. Although beneficial in correcting dyslipidemia, the combinations of statins with fibrates or niacins have the potential for interactions that increase the risks of adverse effects, such as myositis and hepatotoxicity [51]. Hence, treatment of lipoprotein dysregulation warrants a thorough examination of lipoprotein profiles of specific individuals.

### IV. CONCLUSION

Dyslipidemia is a central feature of the metabolic syndrome characterised by increased levels of LDL cholesterol, reduced HDL-cholesterol and elevated plasma triglycerides. The prevalence of the metabolic syndrome is increasing at an exponential rate in developed countries as a result of sedentary lifestyles. This is a crippling problem as lipoprotein dysregulation is significantly associated with increased risk to cardiovascular disease, the leading cause of mortality in the world today.

As researchers continue to grapple with the depth and breadth of information within the literature, Lipoprotein Ontology attempts to delineate the problem into five key areas which cover the classification of lipoproteins, pathways of lipoprotein metabolism, pathophysiology of lipoproteins, causes of lipoprotein dysregulation as well as treatment of dyslipidemia. Not only will this provide a more efficient way of retrieving relevant information for the intended users, this technology will also support the querying of heterogeneous information sources from the different aspects of lipoprotein research to achieve a common aim. The end result would be a complete repository of lipoprotein knowledge and research in one program for the perusal of physicians, researchers and patients.

**REFERENCES**


