Original article

A novel computational drug repurposing approach for Systemic Lupus Erythematosus (SLE) treatment using Semantic Web technologies

Adeeb Noor a, Abdullah Assiri b,∗

a Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 80221, Saudi Arabia
b Department of Clinical Pharmacy, College of Pharmacy, King Khalid University, Abha 62529, Saudi Arabia

1. Introduction

Development of new drugs is costly, time-consuming, and achieves only a minor success rate (Morgan et al. 2011; Kesselheim, Avorn, and Sarpatwari 2016). The traditional approach could potentially take up to 15 years to complete clinical testing and gain authority approval. Furthermore, the failure rate of new drugs is about 95%, and each one that makes it to final approval has an estimated cost of $1 billion (Morgan et al. 2011; Kesselheim, Avorn, and Sarpatwari 2016; Collins 2011). Therefore, there is an urgent need to both reduce costs and accelerate the drug development process while also increasing the success rate for newly-developed drugs. As such, a new approach is vital. This can begin with methods for re-inventing existing drugs by screening them for new indications for either common or rare diseases. ‘Drug repositioning’ or ‘drug repurposing’ has emerged as a prospective approach for helping the bio-pharmaceutical industry, with a primary focus on improving the business of pharmaceuticals and a secondary focus on improving the therapeutic aspect of the industry (Sardana et al. 2011; Ashburn and Thor 2004). Conceptually, drug repurposing is the process of finding and developing new uses for pre-existing drugs, applying them to functions other than the diseases they were originally geared towards (Ashburn and Thor 2004). This strategy relies on having approved compounds with well-characterized pharmacology, which are thus already associated with a safety profile, allowing the time frame for approval to be substantially reduced relative to novel drug development. It is therefore not surprising that among new medications in 2009 (including vaccines and new formulations), 51 were repositioned drugs, and those represented 30% of all approved drugs brought to market that year (Sardana et al. 2011; Graul and Cruces 2011). Moreover, the profits realized from drug repurposing could exceed billions. For instance, an initial attempt in repurposing thalidomide has led to it now being used for multiple myeloma, reaching profits of $272 million in 2003 (Singhal et al. 1999); it also has recently been approved by the United States as a treatment for Type 2 Diabetes due to having dopamine agonist properties similar to bromocriptine (Pijl et al. 2000).

Even though drug repurposing has been highly successful to date, the potential of this strategy is still not fully realized due to a lack of meaningful and complete results (Abdelhakim et al. 2020). This can be attributed to the fact that much of the established knowledge on which drug repurposing depends was obtained through a serendipitous approach rather than a methodical and systematic approach (Mithun and Khairnar Shubham 2020). In addition, the lack of knowledge dissemination between both computer applications and drug data sectors only continues to hinder the progress and expansion of drug development/repurposing (Alshahran and Hoehndorf 2018). Recently, scientists have developed many applications to help elucidate and understand the complexities of biological and pharmacological systems, which are beneficial in this regard (Althubaiti et al. 2019; Assiri and Noor 2020; Noor et al. 2017; Alshahran and Hoehndorf 2018; Campillos et al. 2008; Qu et al. 2009). However, despite these advances, extant computational repurposing studies have focused only on repurposing a single potential drug at a time.

In this work, we draw upon the power of Semantic Web (SW) technologies to mine the interconnections in diverse biomedical data, building an inferential query Semantic Web-based model that is able to identify multiple potential drug candidates for repurposing in the context of a given disease. This approach allows knowledge discovery to distill insights from the complex mechanistic concoction underlying drug-disease entities with an accuracy that ultimately protects patients and their health. The core hypothesis of our strategy is that upon identification of the disease-associated genes that cause dysfunction in disease-related biological pathways and processes, it is possible to identify drugs that act upon the same pathways and processes, thus narrowing the range of potential available drugs to the most selective and specific candidates. Furthermore, this new framework for discriminatory drug discovery has the advantage of requiring the drug to also act upon the single nucleotide polymorphisms (SNPs) identified and described as associated with the disease. To demonstrate the
useability and validity of the computational biology approach for drug repurposing, a test was done on the complex disease Systemic Lupus Erythematosus (SLE), no cure for which is yet available.

2. Materials and methods

This study proposes an integrative drug repurposing framework using SW technologies to precisely enable the discovery of the hidden mechanisms underlying drug-disease interactions. To realize this approach, it was necessary to first develop a knowledge framework that incorporates a variety of data sources, including pharmacological, bimolecular, phenomenological, and genetic data; there is, unfortunately, no single source that serves the purpose. Thus, this data was semantically integrated and stored in a drug repurposing network using the Java framework Jena.\(^{1}\) Carroll et al. (2004) Fig. 1 shows the overall architecture of the drug repurposing framework.

2.1. Data sources

To generate a network of drug-disease associations, five trusted biomedical sources that contain information about drug and disease mechanisms were used. The Pharmacogenomics Knowledge Base (PharmGKB)\(^{1}\),\(^{2}\) Klein et al. (2001) from which we obtained drug-SNPs and disease-SNPs associations, was downloaded in January 2020 (https://www.pharmgkb.org/downloads). The other four sources were all part of the Unified Medical Language System (UMLS) terminology system,\(^{2}\) Bodenreider (2004) and so were extracted from the UMLS (Version2020). These comprised the National Drug File – Reference Terminology (NDF-RT)\(^{2}\), Brown et al. (2004) for drug-indication associations; the National Cancer Institute Thesaurus (NCI)\(^{2}\), de Coronado et al. (2004) and Entrez Gene from the National Center for Biotechnology Information (NCBI)\(^{2}\), Maglott et al. (2005) for disease-gene and gene-pathway associations; and the Gene Ontology (GO)\(^{1}\),\(^{3}\) Ashburner et al. (2000) for gene-biological process associations. These data sources encompassed 1810 drugs and 3062 diseases.

2.2. Constructing the drug repurposing knowledgebase

We used the UMLS as the backbone of the drug repurposing knowledgebase as it was primarily developed to provide an integration system based on terminology similarity in the biomedical domain. Construction of the knowledgebase consisted of two main steps: 1) building a drug-disease semantic network with the Resource Description Framework (RDF) and 2) adding the semantic relationships. Fig. 2 illustrates this process.

2.3. Building the RDF network

UMLS Metathesaurus version 2020 was downloaded from the UMLS knowledge server with a MySQL loading script. Then, all biomedical datasets were either extracted or integrated to build the drug-disease network. The Concept Unique Identifiers (CUIs) provided by the UMLS were used as the main identifiers for drug-disease RDF nodes. Specifically, we used only three files from the UMLS Metathesaurus: MRCONSO for names, MRREL for semantic relations, and MRSTY for semantic type information. We then retrieved from those tables the NCBI, NIH, NDF-RT, and GO ontology datasets using the UMLS CUI IDs, and integrated PharmGKB into the drug-disease network through the x-reference identifier since a reference to the UMLS CUI ID was provided in the PharmGKB dataset. Finally, a Java script was written to generate RDF nodes using the Jena application programming interface (API).

2.4. Normalization and assertion of semantic relationships

Semantic relationships between instances (CUIs) in the RDF network were asserted from the UMLS Metathesaurus and PharmGKB; more specifically, we used a MySQL to query the semantic relationships from MRREL, which is a table in the UMLS Metathesaurus that stores semantic relations, and directly added genetic associations for the drug-disease network from the PharmGKB dataset. Moreover, all semantic relationships between entities have been reviewed and grouped as much as possible. For instance, we grouped four different relations of drug-disease pairs from the NDFRT ontology, may_treat, may_prevent, may_diagnose, and induces, into the single semantic relationship has_indication. We believe that such groupings ensure the simplicity of our knowledge base and remove any redundancy. The ultimate result was a set of ten asserted semantic relationships that linked instances in the knowledgebase. Those relationships were then used to make axioms with which to infer potential drug repurposing candidates.

2.5. Drug repurposing discovery through mining semantic infrastructure

Using description logic and the SW, we defined axioms to identify possible candidates for drug repurposing. We used the SPARQL Query Language for RDF (Apache Jena - Reasoners and rule engines: Jena inference support [Internet]. [cited 2020 Jan 13]. Available from: https://www.w3.org/TR/rdf-sparql-query/) as an inference tool to identify candidate drugs over the drug-disease network under the hypothesis that upon identification of the disease-associated genes that cause dysfunction in disease-related biological pathways and processes, it is possible to identify drugs

---

\(^{1}\) Carroll et al. (2004)

\(^{2}\) Bodenreider (2004)

\(^{3}\) Ashburner et al. (2000)

---

**Fig. 1. Drug repurposing workflow.** To build the drug repurposing knowledgebase, five different data sources representing drug-disease information were integrated using SW and UMLS. Predictions were made using a complex semantic inference query, and the results were validated by literature review.
that also act upon the same pathways and processes and so to identify candidate treatments. We further required that the drug also act upon the same SNPs as identified and described in the disease. Taking the rare disease SLE as a test case, we thus considered multiple biological features and identified drugs that could potentially treat SLE. More specifically, our method starts with a disease as input, and the inference engine then runs over drug-disease networks to find drugs that satisfy predefined biological rules with respect to the targeted disease. Fig. 3 illustrates the complex inference query for this case study.

3. Results

Applied to 1810 FDA-approved drugs, the computational biology approach yielded 11 drugs that satisfy the axioms (i.e. shown to have links with SLE through genes acting in the same pathways and processes, and additionally through acting upon SNPs identified and described as associated with the disease). These drugs could potentially affect 13 genes that may relate to the pathophysiology of SLE through four different biological pathways.

3.1. Gene-pathway associations in relation to SLE

The drug-disease association network yielded four pathways with potential contribution to SLE pathophysiology, and about 13 coding genes as essential elements within them. The identified pathways were: Inflammatory Response Pathway (IL1A, IL5, IL6, IL12, IL17, Lymphotoxin A pathway signaling), T-Cell Polarization Pathway, Hematopoiesis Pathway, and Telomere Pathway. In addition, the genes associated with these pathways were: Cluster of Differentiation 4 (CD4), Interleukin-1A (IL1A), Interleukin-5 (IL5), Interleukin-6 (IL6), Interleukin-17 (IL17), Lymphotixin Alpha (LTA), C-C Motif Chemokine (CCR1), Janus Kinase 2 (JAK2), X-ray Repair Cross-complementing Protein 5 (XRCC5), X-ray Repair Cross-complementing Protein 6 (XRCC6), Retinoblastoma Protein 1 (RB1), MYC Proto-oncogene (MYC), and Tumor Protein 53 (TP53). All identified pathways and the associated genes are summarized in Table 1.

3.2. Drugs candidates associated with SLE

Of the 11 drugs identified by the drug-disease association network ten (91%) are known to have either potential benefit or potential worse outcome in association with SLE. Among these, seven drugs (aspirin, azathioprine, cyclophosphamide, indomethacin, methotrexate, leflunomide, and warfarin) [Iudici et al. 2016; Fanouriakis et al. 2019; Cao et al. 2015] are currently included in the SLE treatment guidelines for one purpose or another, while three drugs (clopidogrel, peginterferon alfa-2a, and peginterferon alfa-2b) [Hewitt, Carton, and Wakelin 2018; Rizvi and Hojjati 2011; Yilmaz and Cimen 2009] are associated with either worsen-
ing outcome or the potential to induce SLE. Only one candidate (propranolol) was not associated with a known benefit or poor outcome. All drugs with associated benefits or harmful outcomes are summarized in Table 2.

4. Discussion

The framework analysis used SLE as a case study in utilizing the SW technologies to identify drugs with potential disease-related effects. This method allows the prediction of biological pathways that potentially contribute to disease progression and subsequently the prediction of drugs having potential association with the identified pathways, which may provide new treatment options or identify medications as having potential harmful effects when used to treat the disease in question.

Among the results of our case study, the Inflammatory Response Pathway was represented by several cytokines and proteins, including Interleukin-17, which is an important cytokine known to be associated with tissue damage in SLE (Nalbandian, Crispín, and Tsokos 2009). Likewise, several cytokines and proteins in the Inflammatory Response Pathway contribute substantially to SLE progression, including interleukin-1 alpha, interleukin-5, interleukin-6, and lymphotoxin-alpha (LT-alpha, LTA) (Wallace et al. 2017; Lieberman and Tsokos 2010; Nalbandian, Crispín, and Tsokos 2009; Tucci et al. 2008; Parks et al. 2004; Wen et al. 2004; Zhang et al. 2015; Hohensinner,

![Table 1](image-url)

**Table 1**

| Gene Interaction | Pathway | Drug Candidate | Evidence |
|------------------|---------|----------------|----------|
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL1A, IL17, LTA) | Aspirin | 24022862 |
| CD4 gene IL1A gene | Inflammatory Response Pathway (IL17) | Azathioprine | NA |
| CD4 gene LTA gene | T-Cell Polarization Pathway | | |
| CD4 gene CCR1 gene | | | |
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL17) | Leflunomide | 11480845 |
| CD4 gene LTA gene | Inflammatory Response Pathway (LTA) | Methotrexate | 9227169 |
| CD4 gene IL6 gene | Hematopoiesis Pathway | | |
| CD4 gene IL6 gene | | | |
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL5, IL6, IL12, IL17) | Cyclophosphamide | 15230294 |
| CD4 gene IL6 gene | | | |
| CD4 gene IL6 gene | | | |
| CD4 gene JAK2 gene | Telomere Pathway | | |
| XRCC5 gene RB1 gene | | | |
| XRCC5 gene MYC gene | | | |
| XRCC6 gene RB1 gene | | | |
| XRCC6 gene MYC gene | | | |
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL1A, IL17) | Indomethacin | 2216092 |
| CD4 gene IL1A gene | T-Cell Polarization Pathway | | |
| CD4 gene CCR1 gene | | | |
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL17) | Leflunomide | 11480845 |
| CD4 gene IL17 wt All | Inflammatory Response Pathway (IL17) | Methotrexate | 9227169 |

(continued on next page)
Regarding the NO2-dependent IL12 Pathway, a study by Tucci et al. found that the overexpression of IL-12 was associated with increased SLE complications, specifically lupus nephritis. Additionally, IL-12 promotes several factors that in turn subsequently promote T-cell polarization and worsening of SLE (Tucci et al. 2008). Lastly, several studies have reported an association for telomere dynamics in SLE patients, which opens a new avenue for therapeutic targeting strategies such as that proposed informatically in this work (Hohensinner, Goronzy, and Weyand 2011).

In addition, the case study yielded 11 drugs, of which ten (91%) were identified as having potential effect on SLE progression through four biological pathways and had prior reports of association with SLE in some respect. For example, aspirin is widely used in SLE patients on account of its potential role in primary prophylaxis of cardiovascular coronary events (Iudici et al. 2016). However, upon further assessment, it was found that a considerable number of SLE patients may have aspirin resistance, suggesting that a treatment alternative be considered for this population (Akdogan et al. 2013). In this work, aspirin was proposed to

| Drug          | Gene Combination | Pathway                                | Protein          | Reference          |
|---------------|------------------|----------------------------------------|------------------|--------------------|
| Propranolol   | CD4 gene LTA gene| Inflammatory Response Pathway (LTA)    | CD4 gene IL.6 gene | Warfarin 10494759 |
| Warfarin      | CD4 gene IL.5 gene| Hematopoiesis Pathway                  | Peginterferon alfa-2a |
| Aspirin       | CD4 gene IL.6 gene| Inflammatory Response Pathway (IL5, IL6, IL12, IL17) | Peginterferon alfa-2b |
| Propranolol   | XRCC5 gene TP53 gene | Telomere Pathway                      | XRCC5 gene TP53 gene |                |
| Warfarin      | XRCC6 gene TP53 gene | Telomere Pathway                      | XRCC6 gene TP53 gene |                |

Goronzy, and Weyand 2011).
contribute to SLE pathophysiology through the Inflammatory Response Pathway, as indicated in Table 1. As with aspirin, azathioprine, cyclophosphamide, methotrexate, and nonsteroidal anti-inflammatory drugs (indomethacin) are all listed in the SLE guidelines to be used with specific recommendations and precautions (Fanouriakis et al. 2019). Both azathioprine and cyclophosphamide are recommended for lupus nephritis and those who have severe SLE, while methotrexate is mainly recommended for arthritis, cutaneous lupus, serositis, and those with severe SLE. In regard to nonsteroidal anti-inflammatory drugs, they are mainly recommended for lupus joint pain (Fanouriakis et al. 2019).

Among the identified drugs, three were associated with unpleasant outcomes in the context of SLE. Recent reports have raised an alert that clopidogrel might trigger SLE, suggesting other alternatives be recommended for patients with such risk (Hewitt, Carton, and Wakelin 2018). Similarly, Yilmaz et al. documented a case in which a patient on peginterferon alfa-2b for treatment of hepatitis B developed SLE, suggesting a potential contribution of this drug to SLE induction. As both peginterferon alfa-2a and peginterferon alfa-2b were identified in our informatic framework, this concern may apply to both forms (Yilmaz and Cimen 2009). Lastly, the framework predicted propranolol to be associated with SLE. Upon reviewing the literature, not a single study was found to explain or provide evidence supporting this potential. However, propranolol may have clinical use in treating SLE-associated symptoms. The potential benefit or risk of propranolol in SLE patients might warrant further investigation if clinical evidence becomes available.

The computational biology approach for drug repurposing undertaken in this study largely yielded findings that had well-supported associations with SLE. However, the pure framework was not able to predict all drugs having established clinical use for SLE treatment, and even for those it did predict, it was not able to rank the candidates. One explanation is that even for drugs currently in clinical use, the exact mechanism of action is not necessarily well-understood, thus key details are not present in current databases. Furthermore, the computational approaches such as the one in this study require the precise and narrow selection of results to minimize false-positive and true-negative findings. Therefore, a broad-spectrum prediction approach might be warranted in future studies, especially for new and rare diseases. Another notable limitation of this work is that the prediction process considered only the existence of an association with SLE, without clear implication as to whether the association is beneficial. Future integration of additional filtering and selection approaches into such a framework would definitely add more value for researchers and clinicians alike.

5. Conclusion

In conclusion, this study utilized advanced technology such as the Semantic Web to predict drugs associated with a rare disease. Taking SLE as a case study, the framework yielded four pathways having potential association with the disease, and ten drugs with the potential to affect SLE either for good or ill. This unique approach opens avenues for predicting new indications for existing drugs while minimizing the potential risk of drug-induced disease. Future studies that draw upon more thoroughly-networked informatic data sources are warranted to help advance drug repurposing and offer new treatment options, especially for new and rare diseases.

Funding

This project was funded by the Deanship of Science Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under grant# (DF-609–611-1441).

Author contributions

AN and AA designed and conducted the study, analyzed and validated the study findings, and wrote the manuscript.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgments

The authors would like to acknowledge with great appreciation the Deanship of Science Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia.

References

Abdelhakim, M., McMurray, E., Syed, A.R., Kafkas, S., Kamau, A.A., Schofield, P.N., Hoehndorf, R., 2020. DDIDM: drug database for inborn errors of metabolism. Orphanet J Rare Dis 15, 146.
Akdogan, A., Kilic, L., Alman, U., Dogan, I., Karadag, O., Bilgen, S.A., Buyukasik, Y., Kiraz, S.; Erenli, L., 2013. Aspirin resistance in systemic lupus erythematosus. A pilot study. Lupus 22, 835–838.
Alshahrani, Mona, and Robert Hoehndorf. 2018. ‘Drug repurposing through joint learning on knowledge graphs and literature’, bioRxiv: 385617.
Althubaiti, S., Karwath, A., Dalil, A., Noor, A., Alkhayyat, S.S., Alwassia, R., Mineta, K., Gojobori, T., Peginterferon alfa-2a & Peginterferon alfa-2b

| Drug Candidate | Better Outcomes | Harmful Outcomes |
|----------------|-----------------|-----------------|
| Aspirin        | Primary prophylaxis of cardiovascular events (Iudici et al. 2016) | NA |
| Azathioprine   | Used for lupus nephritis and severe SLE (Fanouriakis et al. 2019) | NA |
| Clopidogrel    | NA | Potential to induce SLE (Hewitt, Carton, and Wakelin 2018) |
| Cyclophosphamide | Used for lupus nephritis, and severe SLE (Cao et al. 2015) | NA |
| Indomethacin   | Used for lupus joint pain (Fanouriakis et al. 2019) | NA |
| Leflunomide    | Used for lupus nephritis (Cao et al. 2015) | NA |
| Methotrexate   | Used for arthritis, cutaneous lupus, serositis, severe SLE (Fanouriakis et al. 2019) | NA |
| Propranolol    | NA | NA |
| Warfarin       | Used for treatment of antiphospholipid syndrome (APS) in the context of SLE (Fanouriakis et al. 2019) | NA |
| Peginterferon alfa-2a & Peginterferon alfa-2b | NA | Induce lupus in patients with Hepatitis B or C (Yilmaz and Cimen 2009) |
