

# Mycophenolate Mofetil Associated Molecular Profiles and Diseases

Konrad Mönks<sup>1</sup>, Andreas Bernthaler<sup>2</sup>, Irmgard Mühlberger<sup>1</sup>, Bernd Mayer<sup>1</sup>, Rainer Oberbauer<sup>2</sup> and Paul Perco<sup>1\*</sup>

<sup>1</sup>emergentec biodevelopment GmbH, Gersthofer Straße 29-31, 1180 Vienna, Austria

<sup>2</sup>Krankenhaus der Elisabethinen Linz and Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria

## Abstract

**Background:** High-throughput Omics technologies aimed at characterizing the molecular profile of diseases together with massive scientific literature on drugs and clinical trials opened the way for matching molecular profiles and drug mode of action in the realm of drug repositioning. We developed a computational analysis workflow for linking molecular targets, drugs, and diseases, and exemplified this approach for the immunosuppressive drug mycophenolate mofetil (MMF).

**Methods and results:** We first established a molecular MMF footprint consisting of deregulated Omics features from two transcriptomics datasets as well as from molecular features associated with MMF based on literature search methods. This footprint, consisting of 170 unique features, was used to identify diseases of relevance to MMF in the scientific literature using Medical Subject Heading (MeSH) terms. A disease enrichment score was calculated for each disease in the MeSH hierarchy, with highly ranked diseases being potentially associated to MMF.

Diseases currently mentioned in clinical trials on MMF were used to validate our approach. The area under the curve was 0.78 when using the disease enrichment scores in order to discriminate between diseases currently in clinical trials and diseases not addressed by MMF with sensitivity and specificity values of 0.38 and 0.96 respectively. Among those diseases in clinical trials showing high scores were kidney diseases, multiple sclerosis, and systemic lupus erythematosus.

**Conclusion:** We identified a significant recovery of drug-associated diseases for the example case of MMF solely utilizing a molecular profile of the drug mode of action. The approach furthermore provided hypotheses on further diseases approachable by the given drug.

**Keywords:** Mycophenolate mofetil; Drug repositioning; Literature mining; Networks; Medical subject headings

## Introduction

De novo drug discovery and development is a challenging endeavor that usually takes between 10 to 20 years from the initial scientific innovation to a viable drug, showing overall success rates well below 10% [1]. Thus, in the best case, medical needs are addressed decades after manifestation. Drug repositioning, i.e. identifying novel indications for customary drugs, or alternative uses for drug candidates which failed prior to registration, may offer valid alternatives for tackling unmet clinical needs [2]. Among the various strategies applied for rational approaches towards drug repositioning the comparative analysis of transcriptional response [3] as well as utilizing drug-protein networks derived from literature and data mining [4] have been applied. In either case, the challenge is to integrate data describing multiple levels: a specific molecular state (e.g. via Omics profiles), data on the disease phenotype, and data on the intervention strategy (e.g. drug and dosage information).

Confronted with this challenge, literature mining offers the unique opportunity to carry out the integration process needed for drug repositioning in one unifying framework, given that all the different data levels are dealt with in scientific publications. Thus, mining the scientific literature is a systematic approach to exploit the high quality information present therein, and offers significant support to drug discovery and drug repositioning [5]. Given the significant growth rate of the academic output a variety of methods assisting researches in keeping pace with this wealth of information have been devised [6]. The challenge of literature mining consists in translating unstructured, human readable free text into structured, machine readable information. Named Entity Recognition (NER) is one technique of particular importance in life science research. NER aims at classifying

textual elements into predefined categories like for example genes, proteins, drugs or diseases [7]. Gene name normalization is successively used in order to link genes to unique identifiers such as NCBI Entrez GeneIDs [8]. Once a set of documents has been annotated using NER followed by gene name normalization, concept co-occurrence can be used to establish relationships among identified entities. As access to full text is typically limited in the scientific context, exploiting the NCBI curated PubMed annotations with Medical Subject Headings (MeSH) becomes more and more important [9]. The MeSH ontology represents a structured vocabulary of approximately 25,000 concepts including a fine-grained coverage of drugs and diseases.

A drug also indexed in the MeSH ontology is mycophenolate mofetil (MMF) mainly used to suppress the immune response after organ transplantation. MMF is a prodrug of mycophenolic acid (MPA), which is a potent and selective inhibitor of inosine-5' monophosphate dehydrogenase (IMPDH) [10]. IMPDH is the rate-limiting enzyme in the de-novo pathway of guanosine synthesis. T- and B-lymphocytes rely on this pathway as their principal source of guanosine in contrast to other cell types that are capable of using the alternative salvage

**\*Corresponding author:** Paul Perco, emergentec biodevelopment GmbH, Gersthofer Straße 29-31, 1180 Vienna, Austria, Tel: +43-1-4034966-10; Fax: +43-1-4034966-19; E-mail: [paul.perco@emergentec.com](mailto:paul.perco@emergentec.com)

**Received** January 11, 2011; **Accepted** January 11, 2011; **Published** January 13, 2011

**Citation:** Mönks K, Bernthaler A, Mühlberger I, Mayer B, Oberbauer R, et al. (2011) Mycophenolate Mofetil Associated Molecular Profiles and Diseases. J Comput Sci Syst Biol 4: 001-006. doi:10.4172/jcsb.1000068

**Copyright:** © 2011 Mönks K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

pathway [12]. MPA therefore exhibits a selective cytostatic effect on lymphocytes. MMF is approved for the treatment of allograft rejection after renal, cardiac and liver transplantation [10]. Further clinical studies on other indications, including certain autoimmune diseases and lymphoma are currently ongoing.

In the present study, we outline a procedure of linking drugs, diseases, and molecular features based on Omics datasets and literature mining approaches with particular focus on the immunosuppressive drug mycophenolate mofetil. Diseases linked to MMF based on molecular features are discussed in the context of currently ongoing clinical trials on MMF.

## Materials and Methods

### Analysis concept

We utilized public domain Omics sources and literature mining for identifying specific molecular features and consecutively diseases associated with MMF. First a set of molecular features (molecular MMF footprint) was derived by combining results from MMF-specific literature mining and two transcriptomics profiling data sets specifically aimed at reflecting the impact of MMF on the level of protein coding gene expression. Based on this molecular MMF footprint a set of diseases was delineated by computing significant enrichment of disease terms associated to publications being linked to these features, again utilizing literature extraction methods. This resulting set of diseases is thus indirectly related to MMF via molecular features being targets or otherwise affected by MMF. Additionally a second disease list was derived by mining information available for published clinical trials utilizing MMF in treatment. Finally both disease lists, namely on the one hand derived on the basis of the molecular MMF footprint, and on the other hand already evaluated in clinical trials were compared for identifying the overlap of diseases for validating the veracity of our approach, but also to identify potential further diseases which, at least from the viewpoint of molecular profiles, might be addressable by MMF.

### Molecular MMF footprint

Molecular features associated with MMF were extracted from two Omics datasets utilizing MMF in-vitro and in-vivo, as well as derived from the scientific literature available on MMF. One MMF associated Omics dataset was retrieved from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) with the GEO dataset ID GSE13922, and the second dataset was retrieved from a compilation of expression data sets found at the connectivity map [12], specifically selecting transcriptomics profiles characterizing the effect of MMF on the cell line MCF-7. In general, identifying data sets specifically characterizing the cellular response to MMF is difficult. Most MMF associated studies available at Gene Expression Omnibus or Array Express are rooted in the transplant context, hence being afflicted with multiple therapeutic regimes involving a variety of drugs. The GEO data set GSE13922 summarizes the effects of MMF on the transcriptome of carotid endarterectomy samples collected from patients with carotid artery stenosis [13]. Given the strong cytostatic effect of MMF on T-cells, a reduction in plaque inflammation and consequently atherosclerosis and cardiovascular events was expected. In this study setting in total 20 patients suffering from carotid artery stenosis were either treated with a single dose of MMF or placebo. Two weeks after treatment start carotid endarterectomy was carried out and plaque specimens were collected for gene expression profiling using Illumina humanRef-8 v2.0 expression BeadArrays. We retrieved the raw data from GEO and identified genes deregulated between the MMF

and the placebo group utilizing a t-test ( $p$ -value  $< 0.05$ ) and the fold-change criterion ( $> 0.6$ ) on the log<sub>2</sub> normalized data.

As second source a subset of samples from the connectivity map (CMAP) was used for deriving differentially expressed features linked to MMF. The connectivity map in its current version (build 2.0) contains more than 7,000 expression profiles representing 1,309 compounds and seeks to document the response of various cell lines to a wide variety of small molecules and drugs, including mycophenolate mofetil [12]. We extracted the raw data (Affymetrix Human Genome U133A platform) of 19 arrays of MMF treated cell lines. Consolidation of technical replicates (mean expression values) led to three biological MMF samples and three control samples for the breast cancer cell line MCF-7. After preprocessing the expression profiles a combination of fold-change ( $> 0.6$ ) and paired t-test ( $p$ -value  $< 0.05$ ) was used to identify deregulated transcripts.

For the literature search also aimed at identifying MMF associated molecular features the Fast Automated Biomedical Literature Extraction (FABLE, <http://fable.chop.edu>) [14] tool combined with a methodology for identifying significantly enriched features was applied. In order to identify genes and proteins that are enriched in biomedical publications related to MMF the occurrence of a gene/protein in an article needs to be verified in the first place, i.e. a mapping between a list of unique identifiers (Ensemble GeneID or Entrez GeneSymbol) and a set of PubMed articles has to be established. Once this mapping is in hand, statistical enrichment analysis can be carried out.

The FABLE algorithm consists of two steps: First, a statistical classifier is used to train a probabilistic model which serves as basis for gene tagging, i.e. for identifying possible occurrences of a gene taking the textual context into account. Given such an occurrence exhibits a sufficient likelihood of actually representing a gene, this occurrence is normalized in a second step to the official Gene Symbol. This normalization step is based on gene synonym lists, which are compared to the predicted occurrence using both exact and relaxed pattern matching procedures. It has been shown that this approach is competitive to alternative methods such as standard information extraction techniques and direct pattern matching both in terms of precision and recall [14,15]. We applied this procedures to all papers retrieved from PubMed associated with “mycophenolate mofetil” (PubMed status as of March 2010).

The second literature analysis approach was based on a query for the MeSH term “mycophenolate mofetil” in the set of publications covered by the mapping from gene-2-pubmed as provided by NCBI (status as of April 2010, <ftp://ftp.ncbi.nih.gov/gene/DATA/gene2pubmed.gz>). The papers retrieved by this query represent the result set, while all the about 270,000 papers represented in the gene2pubmed.gz file depict the background population. In both sets the frequencies of MMF associated genes were determined, i.e. the number of articles mentioning a particular gene, and thus a Fisher's exact test could be applied to determine the level of enrichment. A normalization step was successively introduced in order to account for papers reporting high numbers of genes for taking the total number of reported genes per publication into account when calculating gene frequencies. Genes with a  $p$ -value below 0.05 were considered as significantly associated to MMF based on the scientific literature.

The molecular MMF footprint was finally constructed by forming the union over the above described data sets, namely constructing a list of unique molecular features taking the two Omics sets (GEO, CMAP) as well as the two literature datasets (FABLE, gene-2-pubmed) into account. The feature overlap was determined and the assignment

of features to molecular pathways using the PANTHER Classification System was performed [16].

### Relating the molecular MMF footprint to diseases

For the genes listed in the molecular MMF footprint the assigned publications were determined again utilizing the NCBI-curated gene-2-pubmed file. Each article in PubMed is annotated with a set of MeSH terms, among them disease terms or disease categories, which in turn allowed the calculation of a disease enrichment score (DEscore) for each disease mentioned within this set of publications: Disease enrichment for individual genes listed in the molecular MMF footprint was calculated based on Fisher's exact tests. Enrichment scores for the four feature list (from the connectivity map CMAP, from the GEO data set GSE13922, from FABLE- or MeSH-based publications analysis) were successively combined into gene-disease matrices. For each feature list a matrix was computed in which each entry represents the enrichment p-value of a certain disease within the set of papers that are associated with a certain feature. In a second step, these gene-specific enrichments represented by p-values resulting from statistical testing were inverted and averaged for each disease yielding a score positively correlated with the relevance of the disease with respect to the given molecular feature set. For each molecular feature list this score was standardized, i.e. the mean was set to zero and the standard deviation was set to one. Scores obtained from the four lists were summed up for each disease thus obtaining the final DEscore. The ranked list of diseases based on the DEscore made up the disease profile of MMF based on the molecular MMF footprint. Diseases with the highest scores were on top of the list being most likely associated to MMF.

Diseases apparently related to MMF were furthermore extracted from information on clinical trials effectively using MMF. We used the database at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) maintained by NIH, retrieved all trials related to MMF based on the search term "mycophenolate mofetil", and extracted the associated disease terms from the accompanying MeSH term.

Finally, the list of diseases derived on the basis of the molecular MMF footprint was ranked based on the DEscore, and information whether the disease was effectively mentioned in at least one clinical trial on MMF was added. Sensitivity and specificity values of predicting whether a certain disease delineated on the basis of the molecular MMF feature list was already mentioned in a clinical trial was evaluated by using the DEscores at different cutoff values. The receiver operator characteristic (ROC) curve was generated and the area under the curve (AUC) was determined.

Diseases identified as adverse events of MMF based on the FDA approved drug label (<http://www.gene.com/gene/products/information/cellcept/>) were flagged and excluded from further analyses.

## Results

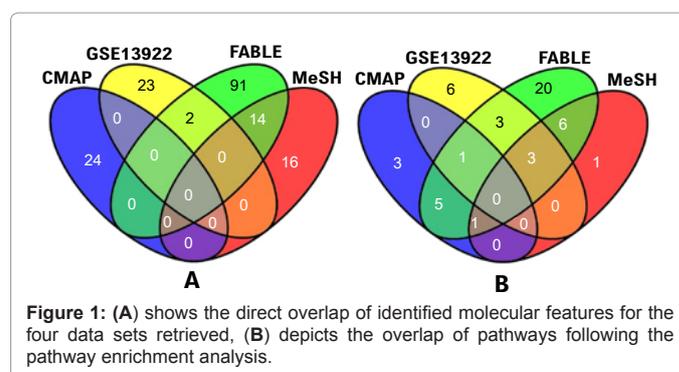
### Molecular MMF footprint

The Omics data set (GEO GSE13922) comparing MMF treated patients suffering from carotid artery stenosis and respective controls identified 25 genes as differentially regulated. Correspondingly, 24 deregulated genes were identified for the MMF treated tumor cell lines compared to untreated controls as provided in the CMAP dataset. 107 genes were identified in the top 10 percent resulting from the FABLE literature mining approach, and the MeSH literature mining approach provided 30 genes with a p-value < 0.05 as being significantly related to MMF. In total 186 molecular features were found to be associated with MMF utilizing the four data sets, representing 170 unique molecular features.

Whereas the feature overlap between the two literature mining approaches was significant with almost 50% of features from the MeSH approach being also in the FABLE feature set, only two of the literature features were also found to be deregulated on the mRNA level in one of the omics datasets as given in Figure 1A.

On the level of molecular pathways the overall picture was different. 19 out of 49 pathways were enriched in features of at least two data sets, with five pathways being of relevance in three of the four data sets (Figure 1B). These five pathways included 'apoptosis signaling', 'EGF receptor signaling', 'inflammation mediated by chemokine and cytokine signaling', 'interleukin signaling' and 'T cell activation'. The pathway enrichment was calculated on the basis of the PANTHER classification system [16].

Table 1 holds those molecular features identified in at least two of the four datasets under study. Among them are the MMF targets IMP dehydrogenase 1 and 2 (IMPDH1 and IMPDH2). The most



| gene symbol | data set short name | gene name  |
|-------------|---------------------|--|
| ABCB1       | FABLE, MeSH         | ATP-binding cassette, sub-family B (MDR/TAP), member 1               |
| ABCC2       | FABLE, MeSH         | ATP-binding cassette, sub-family C (CFTR/MRP), member 2              |
| C3          | FABLE, GEO          | complement component 3   |
| CD55        | FABLE, MeSH         | CD55 molecule, decay accelerating factor for complement              |
| CRP         | FABLE, MeSH         | C-reactive protein, pentraxin-related                                |
| CYP3A5      | FABLE, MeSH         | cytochrome P450, family 3, subfamily A, polypeptide 5                |
| ICAM1       | FABLE, MeSH         | intercellular adhesion molecule 1                                    |
| IMPDH1      | FABLE, MeSH         | IMP (inosine 5'-monophosphate) dehydrogenase 1                       |
| IMPDH2      | FABLE, MeSH         | IMP (inosine 5'-monophosphate) dehydrogenase 2                       |
| IL10        | FABLE, MeSH         | interleukin 10   |
| IL6         | FABLE, GEO          | interleukin 6 (interferon, beta 2)                                   |
| NFKB1       | FABLE, MeSH         | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 |
| TGFB1       | FABLE, MeSH         | transforming growth factor, beta 1                                   |
| UGT1A8      | FABLE, MeSH         | UDP glucuronosyltransferase 1 family, polypeptide A8                 |
| UGT1A9      | FABLE, MeSH         | UDP glucuronosyltransferase 1 family, polypeptide A9                 |
| UGT2B7      | FABLE, MeSH         | UDP glucuronosyltransferase 2 family, polypeptide B7                 |

Table 1: Provided are the Gene Symbol, the data set short name, and gene name of MMF associated molecular features identified in at least two out of four datasets.

enriched Gene Ontology biological processes for these 16 features were ‘regulation of immune response’ (GO:0050776, p-value = 4.24e-8) and ‘inflammatory response’ (GO:0006954, p-value = 3.66e-7).

### MMF disease ranking

For the in total 170 unique molecular features assigned to MMF associated diseases based on publication assignment were retrieved and statistically evaluated, leading to a disease enrichment score (DEscore). The distribution of DEscores, as depicted in Figure 2A, is given for the different MeSH hierarchy levels with the general hierarchy levels one and two showing much higher scores on average. MeSH terms, including disease concepts, are organized in a directed acyclic graph, which can also be represented as hierarchical structure with multiple occurring terms. In this structure, higher level terms (superterms) represent more general concepts while low level terms (subterms) are rather specific disease descriptions. Each subterm may belong to multiple superterms. For example, the term ‘HIV infections’ is a subterm of both superterms ‘Virus Diseases’ and ‘Immune System Disease’.

Diseases already associated with MMF in the context of clinical trials received higher scores than those diseases currently not found in clinical trials using MMF. Figure 2B shows the number of enriched diseases per MeSH hierarchy level. The distribution largely resembles the overall distribution of disease terms in the MeSH hierarchy with most disease terms encountered on the intermediate levels of the hierarchy (levels 3 to 6). Very generic disease terms on levels one and two, and very specific disease terms on levels seven and eight are less frequent.

Based on the DEscore a prediction was performed whether diseases found in clinical trials utilizing MMF can be identified based on the diseases associated with the molecular MMF footprint. Performing such a prediction resulted in an area under the curve of the resulting ROC curve of 0.78, as shown in Figure 3. At a DEscore cutoff of 2.5, sensitivity and specificity values were 0.38 and 0.96 respectively. This cutoff was used in order to identify highly ranked diseases already in clinical trials, but also to identify further diseases of potential relevance in the context of MMF.

Based on the FDA approved drug label a list of adverse events of MMF was manually curated. In a second step, this list was mapped to the diseases identified by our DEscore as relevant to MMF. About 60% of these diseases were identified as being listed as adverse events of MMF and thus excluded from further analysis. 29 diseases on

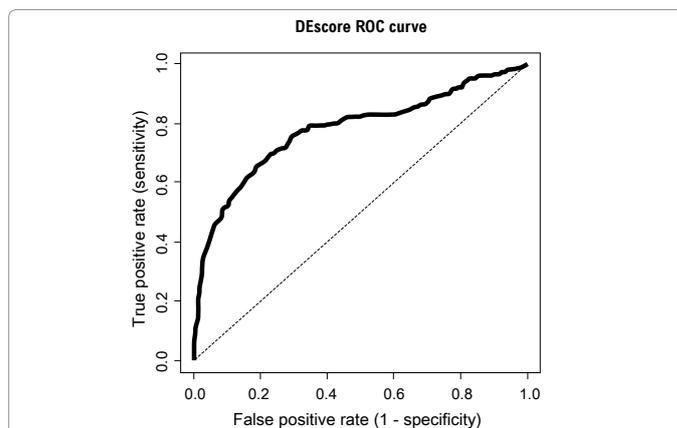


Figure 3: ROC curve of the DEscore for predicting the disease occurrence in clinical trials using MMF (AUC = 0.78).

| disease term                           | DEscore |
|--|---------|
| HIV infections                         | 11.46   |
| kidney diseases                        | 9.19    |
| renal insufficiency                    | 7.68    |
| liver cirrhosis                        | 7.64    |
| gastroenteritis                        | 7.26    |
| demyelinating autoimmune diseases, CNS | 7.19    |
| multiple sclerosis                     | 6.92    |
| hepatitis                              | 6.21    |
| arthritis                              | 6.17    |
| breast diseases                        | 5.65    |

Table 2: Highly ranked diseases currently in clinical trials utilizing MMF.

MeSH hierarchy levels three to five remained with DEscores above 2.5 currently in clinical trials on MMF with the top 10 given in Table 2.

### Discussion

The molecular MMF footprint presented in this work was derived by combining Omics data sets and literature mining results. Clearly evident is the lack of overlap on the individual feature level for the two Omics data sets. Next to the frequent finding of weak feature overlap even for homogeneous Omics studies the given data sets rest on highly heterogeneous sample material, being a cancer cell line on the one hand, and human samples in the realm of cardiovascular disease on the other hand. In contrast to the weak Omics feature overlap, the overlap of literature features is significant, indicating the consistency of the chosen literature mining methods (Fable and MeSH). Those feature found in more than one feature set are, in general, of direct relevance to inflammatory processes (IL6, NFKB1), are related to immune response (CRP, CD55, ICAM1, C3) or are involved in drug resistance (ABCB1, ABCC2). Not surprisingly, the primary target of MMF (IMPDH1, IMPDH2) is also recovered. The overlap on a functional level based on enriched pathways and biological processes was significant. Of the 19 commonly found pathways the top five were all related to immune response. The functional characterization of the most relevant features that were found in at least two datasets also resulted in the GO biological process terms “regulation of immune response” and “inflammatory response” as the top ranked categories. Apparently on a functional level the mode of action of MMF is correctly mirrored by the molecular MMF footprint.

For the link between molecular features and diseases we decided to use MeSH term assignments as well as gene-2-pubmed relations

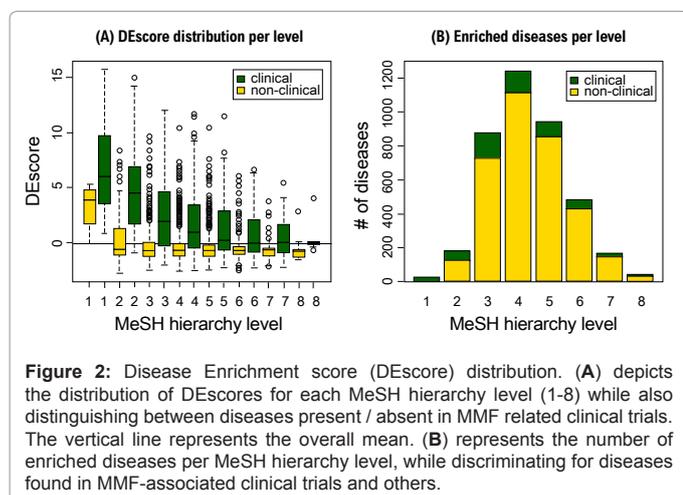


Figure 2: Disease Enrichment score (DEscore) distribution. (A) depicts the distribution of DEscores for each MeSH hierarchy level (1-8) while also distinguishing between diseases present / absent in MMF related clinical trials. The vertical line represents the overall mean. (B) represents the number of enriched diseases per MeSH hierarchy level, while discriminating for diseases found in MMF-associated clinical trials and others.

provided by NCBI. Both mappings, namely genes to articles and MeSH terms to articles represent high quality annotations with the further benefit of the clearly defined structure in the MeSH ontology. This ontology allows for an easy mapping of diseases to scientific articles as well as clinical trials [17].

In the calculation of the DEscore we had to account for two issues, namely (i) the degree of annotation of single genes in the scientific literature and (ii) the different sizes of molecular feature lists. Some genes are well studied, while other genes are only mentioned in a handful of publications so far. This fact was also described and quantified in a publication by Kemmer and colleagues reporting the Gene Characterization Index [18]. As a consequence, when retrieving a set of articles related to a set of genes, the article set is likely to be biased towards the highly discussed genes. Thus, when identifying enriched diseases within this set, they will not reflect the whole set of features, but rather the well-studied genes and proteins. We therefore extracted sets of articles related to individual genes instead of sets of genes, and identified enriched diseases for each article set. To obtain a measure for the relevance of a disease with respect to a whole set of features, we averaged over the individual gene-specific enrichment scores.

The feature lists upon which the consensus feature set is based are of different sizes. More related articles will naturally be identified when querying PubMed with a larger set of features as was the case for our FABLE literature list holding 172 features as compared to the two Omics lists holding 24 and 25 features respectively. Thus, when retrieving a set of articles related to the consensus feature set, this article set would be biased towards FABLE, and consequently the diseases within this set would rather reflect the FABLE features than the consensus feature set as a whole. We therefore handled each feature list individually thus obtaining four different enrichment scores for each disease. To assure equal contributions from each list the obtained scores were standardized before they were summed up to obtain the final Disease Enrichment score (DEscore).

One result provided by the analysis of the calculated DEscore distribution per ontology level is that rather generic disease terms were typically of higher significance than more specific ones. Given that the structure of the ontology was taken into account for significance calculation, this shift towards general disease concepts was not a methodological bias but rather indicated that MMF was frequently discussed in a non-specific disease context, e.g. kidney diseases. Across all levels of the MeSH ontology, diseases currently under investigation in associated MMF clinical trials generally obtained higher DEscores than diseases currently not mentioned in MMF clinical trials. This strong correlation between our DEscore and the presence of an indication in clinical trials was also reflected in the obtained AUC of 0.78.

After thresholding the DEscore the remaining diseases were split into known indications and novel diseases of relevance to MMF. Next to kidney diseases, in part reflecting end stage renal disease and thus also MMF treatment in the context of renal transplantation, we identified systemic lupus erythematosus, multiple sclerosis, and arthritis among the top ranked diseases based on our DEscore. The top ranked category HIV may result from a co-occurrence in the transplant setting, on the other hand is the combined impact of MMF and highly active antiretroviral therapy using abacavir or efavirenz currently under investigation [19] (clinical trial NCT00021489).

Among those diseases showing the highest DEscores currently not in clinical trials on MMF are Inflammatory Bowel Disease (IBD), Crohn's disease, and Celiac disease. For IBD a series of steroid-based therapies are commonly used whereas in Crohn's disease the anti-

inflammatory drug mesalazine is one of the treatment options. Apart from these intestinal anti-inflammatory agents, immunomodulatory regimes based on azathioprine, prednisone or tacrolimus are frequently applied. The use of MMF, in contrast, has only been considered so far for patients who are refractory or intolerant to conventional first-line agents such as mesalazine steroids [20]. The viability of MMF in this context is controversially discussed. In specific cases it has been described as safe, well tolerated and efficient [21] while others consider it to be inappropriate [22]. From the molecular feature set addressed by MMF such therapeutic approaches seem reasonable although the side effects may outweigh the benefits of the treatment. Further investigations pursuing a personalized approach are necessary to identify the exact circumstances under which a beneficial application of MMF is possible.

The fact that we currently cannot automatically distinguish whether a disease with a high DEscore could be a novel indication for MMF treatment or whether it is merely associated or even an adverse event description is a limitation of our approach. With the ranked disease list in hand clinical expert knowledge is needed for result interpretation. The challenge to be met consists in automatically distinguishing different types of co-occurrences of drugs, diseases and molecular features. Indeed, there are methods for this exact type of task, known as "relations mining" [23,24], but all of them suffer from elevated false-positive and false-negative rates. Our current approach may well serve as starting point for more sophisticated literature mining approaches in the context of drug repositioning.

#### Acknowledgments

Financial support for this study was obtained from Roche Austria GmbH.

#### References

1. Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3: 673-683.
2. Chong CR, Sullivan DJ (2007) New uses for old drugs. *Nature* 448: 645-646.
3. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 107: 14621-14626.
4. Von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R (2010) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res* 39: 1060-1066.
5. Plake C, Schroeder M (2010) Computational Polypharmacology with Text Mining and Ontologies. *Curr Pharm Biotechnol* 12: 449-57.
6. Harmston N, Filsell W, Stumpf MPH (2010) What the papers say: text mining for genomics and systems biology. *Hum Genomics* 5: 17-29.
7. Cohen AM, Hersh WR (2005) A survey of current work in biomedical text mining. *Brief Bioinform* 6: 57-71.
8. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, et al. (2008) Overview of BioCreative II gene normalization. *Genome Biol* 9: S3.
9. Nakazato T, Bono H, Matsuda H, Takagi T (2009) Gendoo: functional profiling of gene and disease features using MeSH vocabulary. *Nucleic Acids Res* 37: W166-169.
10. Villarreal MC, Hidalgo M, Jimeno A (2009) Mycophenolate mofetil: An update. *Drugs Today* 45: 521-532.
11. Allison AC (2005) Mechanisms of action of mycophenolate mofetil. *Lupus* 14: s2-8.
12. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929-1935.
13. van Leuven SI, van Wijk DF, Volger OL, de Vries JP, van der Loos CM, et al. (2010) Mycophenolate mofetil attenuates plaque inflammation in patients with symptomatic carotid artery stenosis. *Atherosclerosis* 211: 231-236.

14. Crim J, McDonald R, Pereira F (2005) Automatically annotating documents with normalized gene lists. *BMC Bioinformatics* 6: S13.
15. McDonald R, Pereira F (2005) Identifying gene and protein mentions in text using conditional random fields. *BMC Bioinformatics* 6: S6.
16. Mi H, Thomas P (2009) PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol* 563: 123-40.
17. Jimeno-Yepes A, Bertanga-Llavori R, Rebholz-Schuhmann D (2009) Exploitation of ontological resources for scientific literature analysis: searching genes and related diseases. *Conf Proc IEEE Eng Med Biol Soc* 2009: 7073-7078.
18. Kemmer D, Podowski RM, Yusuf D, Brumm J, Cheung W, et al. (2008) Gene characterization index: assessing the depth of gene annotation. *PLoS ONE* 3: e1440.
19. Millán O, Brunet M, Martorell J, García F, Vidal E, et al. (2005) Pharmacokinetics and pharmacodynamics of low dose mycophenolate mofetil in HIV-infected patients treated with abacavir, efavirenz and nelfinavir. *Clin Pharmacokinet* 44: 525-538.
20. van Dieren JM, Kuipers EJ, Samsom JN, Nieuwenhuis EE, van der Woude CJ (2006) Revisiting the immunomodulators tacrolimus, methotrexate, and mycophenolate mofetil: their mechanisms of action and role in the treatment of IBD. *Inflamm. Bowel Dis* 12: 311-327.
21. Tan T, Lawrance IC (2009) Use of mycophenolate mofetil in inflammatory bowel disease. *World J Gastroenterol* 15: 1594-1599.
22. Caprilli R, Angelucci E, Cocco A, Viscido A, Annese V, et al. (2005) Appropriateness of immunosuppressive drugs in inflammatory bowel diseases assessed by RAND method: Italian Group for IBD (IG-IBD) position statement. *Dig Liver Dis* 37: 407-417.
23. Ohta T, Pyysalo S, Kim J, Tsujii J (2010) A re-evaluation of biomedical named entity-term relations. *J Bioinform Comput Biol* 8: 917-928.
24. Naeem H, Küffner R, Csaba G, Zimmer R (2010) miRSeI: automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinformatics* 11: 135.