Predicting differential drug response using demographic-based variation in gene expression profiles

Michael Januszyk and Sarah Xing

INTRODUCTION

Demographic factors have frequently been shown to influence disease outcomes and drug response. However, these relationships are typically derived through de facto trial and error in the clinic, and the associated learning curve is expensive both in terms of healthcare costs and patient safety[1]. The need for a systematic strategy to prospectively identify such interactions is well-recognized within the medical community[1,2]; however, until recently, the biomedical informatics infrastructure has been insufficient to support such large-scale endeavors, and these goals remain largely unattained[3].

Clinical decision support models using known pharmacologic relationships have proven highly effective for identifying and avoiding specific adverse drug-drug interactions[4]. They have also been used successfully to modulate dosing for a handful of high-risk drugs using demographic and genotypic information[5]. However, due in large part to inconsistencies within the structure of electronic medical records (EMR), the overwhelming majority fail to routinely incorporate even the most basic non-pharmaceutical information[6].

EMR also provide a fertile testing ground on which to rapidly evaluate suspected clinical relationships. EMR-based approaches have inherent advantages over traditional retrospective chart reviews in terms of both efficiency and accuracy[6]. This form of "in silico" validation is a particularly powerful tool when paired with public data-driven approaches to biomedical discovery, as this effectively affords an environment of paperless, zero patient-risk clinical research.

We developed a novel bioinformatics pipeline with which to identify candidate drugs for which specific demographic attributes may modulate clinical outcomes. In order to evaluate putative drug-demographic relationships, we apply various machine learning techniques using high temporal resolution patient data obtained from an electronic health records database. Our goal is to leverage genomics-derived knowledge in order to more effectively identify patients likely to elicit an abnormal response to a given drug. Such a method would provide a useful adjuvant to the existing paradigm of drug-based decision support models.

METHODS

We constructed a biomedical informatic pipeline using publically available genomics and pharmacological databases in order to identify putative drug-demographic associations. This consisted of three principal phases (Figure 1): (1) Identify genes whose differential expression profiles exhibit demographic-linkage over the scope of thousands of microarray experiments. (2) Translate these demographic-gene associations into demographic-gene-drug or demographic-gene-protein-drug relationships using pharmacologic databases. (3) Evaluate putative demographic-drug relationships through the construction of machine learning models trained using an EMR database.

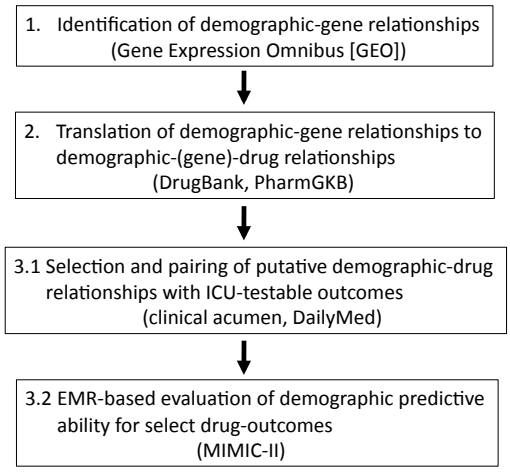

Fig. 1. Bioinformatic pipeline for the identification and evaluation of putative demographic-drug associations.

1. Identifying a list of candidate genes

We began with a meta analysis of the 3,500 human microarray datasets currently available in GEO (Gene Expression Omnibus), a public genomics data repository maintained by the NIH[7]. Our initial query for gender-specific human microarray series produced 91 datasets with at least one sample from each sex, corresponding to 6,714

individual microarray samples (mean = 74.6 chips/dataset) and spanning four distinct GPL platforms. Individual microarray chips were classified according to gender, as well as one principal experimental variable (typically subject disease status [e.g., healthy vs. diseased] or tissue of origin [e.g., bone marrow vs. blood]), and each dataset partitioned accordingly.

For each dataset ($j \in J$) we computed separate mean signal intensity values (proportional to the logarithm of mRNA content) for samples belonging to each of these four groups (male treatment, male control, female treatment, and female control), as well as the associated group-wide variance. We then evaluated the differences in mean intensities for male (treatment vs. control) and female (treatment vs. control) samples,

$$\Delta I^{(j)}_{Male} = \bar{I}^{(j)}_{Male,Treatment} - \bar{I}^{(j)}_{Male,Control}$$

$$\Delta I^{(j)}_{Female} = \bar{I}^{(j)}_{Female,Treatment} - \bar{I}^{(j)}_{Female,Control}$$

and defined the gender-based differential response as the difference of these differences ($\Delta\Delta I = \Delta I_{Male} - \Delta I_{Female}$), which is the log ratio of mRNA fold-change variation in response to treatment.

We computed gender-based differential responses ($\Delta\Delta I$'s) independently for every unique gene (~25,000) within each dataset. We then combined the $\Delta\Delta I$ values from each dataset *j* according to:

$$\overline{\Delta\Delta I} = \sum_{j \in J} \Delta\Delta I^{(j)} \bullet n^{(j)} / (\sigma^{(j)})^2$$

That is, for every gene, the mean change in signal intensity response associated with each dataset was weighted according to both the number of samples in that dataset ($n^{(j)}$), as well as the variance among those samples ($[\sigma^{(j)}]^2$), in order to arrive at a single dataset-wide mean change in signal intensity response ($[\Delta\Delta I]_{bar}$).

2. Classifying and ranking candidate drugs

PharmGKB and DrugBank are publicly available, human-curated pharmacology databases containing, among other parameters, confirmed or suspected drug-gene interactions. All 24,834 genes obtained from the GEO analysis described above were routed through PharmGKB and DrugBank to identify known drug-gene interactions overlapping with our top $\Delta\Delta I$-rated genes. Gene-drug interactions (and, implicitly, gene-protein-drug interactions) were stratified according to the nature of interactions (e.g., target, transporter, and enzyme) and cross-listed between databases using DrugBank IDs. Weighted scores were then constructed for each drug according to the average $\Delta\Delta I$ associated with each of its gene interactions divided by the total number of gene interactions for the given drug. Non-FDA approved drugs were excluded from this analysis. Thus, a list of drugs with suspected drug-gene interactions was produced and sorted by weighted-$\Delta\Delta I$; because $\Delta\Delta I$ is a proxy for demographics factors (gender, in this case), the list can also be interpreted as being sorted by predicted degree of drug-gender interaction.

3. Evaluation of putative drug-demographic relationships using electronic health records

MIMIC II (Multiparameter Intelligent Monitoring in Intensive Care) is a publicly available repository of electronic medical records data (including lab values) from the Beth Israel Deaconess Medical Center intensive care unit covering approximately 26,000 patients across multiple ICU admissions, and also contains high temporal resolution telemetry data for a subset of patients. Telemetry data includes second-by-second measures of respiration rate, heart rate, arterial systolic blood pressure, arterial diastolic blood pressure, and mean arterial blood pressure. Using MIMIC, putative demographics-driven variations in drug response can be evaluated for top candidate drugs identified from the steps above.

For each top-ranking drug in our list, we identified the number of MIMIC patients to which it was administered, as well as the frequency of administration. As expected, nearly half of the drugs in our list were never administered to patients in MIMIC, consistent with the character of an ICU population. The remaining drugs, along with their relative ranking and administration frequency, were manually reviewed by a physician (M.J.) in order to identify those whose efficacy was most readily testable given the breadth and temporal resolution of the parameters available in the MIMIC ICU database. Testable outcomes were limited to lab values (e.g., serum Na^+, K^+) and telemetry (including cardinal vital signs). Using these data, we were able to assess how each patient's values responded to a target drug administration in order to study the pharmacodynamic effects of the drug on the patient's system.

We identified the antiarrhythmic drug *Amiodarone*, the antihypertensive drug *Furosemide*,

and the heart failure drug *Metoprolol*, as frequently administered drugs with testable outcomes seated high on our drug-(gene)-demographic interaction list. Amiodarone and Metoprolol are known to decrease heart rate (sometimes to the point of bradycardia), and Furosemide and Metoprolol both decrease systolic blood pressure (sometimes to the point of frank hypotension). We evaluated whether demographic factors will have a differential effect size on modulation of heart rate and blood pressure.

To test this hypothesis, we queried MIMIC II for all available data on patients receiving these three drugs in the ICU. For every patient, the first pharmacy order for each drug was identified, and an administration schedule derived to determine the earliest administration date and time. A similar process was repeated for each pharmacy order for each patient to identify any other drugs administered within 24 hours of this time point. Demographics data (age, sex) and route of administration were also determined.

Of the patients initially identified in MIMIC as having received each drug, roughly half had corresponding telemetry data overlapping the drug administration period. For each patient, we averaged telemetry data over every 60 second period and identified the minimum heart rate and blood pressure within a one hour window following the first administration of each drug. We then classified patients as having a slow heart rate (determined as minimum heart rate less than 70) or low blood pressure (determined as minimum systolic blood pressure less than 90) during the one hour window following first administration – such patients appear to be strong responders, sometimes pathologically so.

We then sought to determine whether demographic factors were significant predictors of differential response to our target drug. For each patient, a set of binarized features was determined: age over 70, sex, oral or IV route of drug administration, and previous administration of ~500 other drugs, each of which was administered to at least one patient in our dataset. We then performed feature selection using the sequential forward search algorithm[8] with ten-fold cross-validation as the optimization function over the following classifiers: (1) Logistic Regression, (2) Random Forest, (3) Support Vector Machine, and (4) Naïve Bayes with Laplace Smoothing. For each classifier, the results of 10-fold cross validation (10-Fold CV) were recorded using strictly pharmaceutical factors, non-pharmaceutical factors (e.g., age, gender), and features selected from the combined set.

RESULTS

Generation of candidate gene list

The top ΔΔI-ranked genes resulting from our initial analysis of GEO are shown in Figure 2. As expected, several of the top genes by ΔΔI score are located on sex-linked chromosomes (X/Y), including CLIC2, USP9Y, and DDX3Y. Interestingly, among the non-sex-linked genes, the matrix metallopeptidase MMP8 ranked highest. This protein is critical to bone remodeling, and these differences may reflect, in part, gender-based differences in bone density and composition.

Gene Symbol	Chromosome	Gene Name
MMP8	11	matrix metallopeptidase 8
CLIC2	X	chloride intracellular channel 2
LCN2	9	lipocalin 2
PVALB	22	parvalbumin
USP9Y	Y	ubiquitin specific peptidase 9, Y-linked
CEACAM6	19	carcinoembryonic antigen-related cell adhesion molecule 6
IDO1	8	indoleamine 2,3-dioxygenase 1
DDX3Y	Y	DEAD box polypeptide 3

Fig. 2. Selection of top ΔΔI-ranked genes from GEO human microarray analysis.

Identification of candidate drugs

We combined our list of top gene candidates with gene-drug interaction data obtained from DrugBank and PharmGKB in order to evaluate whether demographic factors associated with differentially responsive genes can be linked via known gene-drug interactions to demographically, differentially responsive drugs. Among the most highly implicated drugs, many had physiologic axes that intersected trivially with gender (e.g., *Estradiol, Testosterone*), and others could not be tested either because that drug was not given routinely in the ICU (e.g., *Melatonin*) or no readily evaluated EMR outcomes were available (e.g., *Tamoxifen)*. Of the remaining top drugs, three were manually selected (M.J.), and paired with one or two testable clinical outcomes (Figure 3).

Drug	Class	Rank	Clinical Outcome
Amiodarone	Antiarrhythmic	20	bradycardia
Metoprolol	Beta-blocker	31	hypotension, bradycardia
Furosemide	Diuretic	58	hypotension

Fig. 3. Top-ranked, testable candidate drugs.

Amiodarone (bradycardia)	**Metoprolol** (bradycardia)	**Metoprolol** (hypotension)	**Furosemide** (hypotension)
Baseline HR	Baseline HR	Route	Metoprolol
Gender	Age	Baseline SBP	Baseline SBP
Metoprolol	Route	Pantoprazole	Atorvastatin
Age	Gender	Gender	Aspirin
Route	Nitroglycerin	Age	Calcium
Furosemide	Acetaminophen	Atorvastatin	Vancomycin
Aspirin	Amiodarone	Furosemide	Amiodarone
Albuterol	Albuterol	Aspirin	Clopidogril

Fig. 4. Top-selected features for each drug-outcome by vote among the four models (LR, RF, SVM, NB).

Machine learning

For each of these candidate drugs, we constructed four classifiers: Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), and Naïve Bayes (NB) to predict the associated clinical response (bradycardia and/or hypotension) as a binary variable, corresponding to strong and weak responders as described above. We initially limited our input features to the list of co-administered drugs (i.e., all drugs given to the patient within a 24 hour window beforehand), which is the set routinely employed by EMR-based decision support models[9]. We then repeated our training evaluation using exclusively non-pharmaceutical factors (e.g., age, gender, relevant vital signs), and performed a third evaluation using the combined feature set.

Due to the large number of features (>500) relative to the number of patients in the pharmaceutical and combined feature sets, we employed feature selection using the sequential forward search algorithm (with ten-fold cross-validation) to determine a set of features for each drug-classifier pair. Figure 4 shows the results (condensed for clarity by voting among the four models) of the top 8 features selected for each drug.

Amiodarone (bradycardia)			
	Pharmaceutical only	Non-pharmaceutical	All features
Logistic Regression	73.5%	73.7%	83.8%
Random Forest	80.8%	83.8%	**83.9%**
SVM	78.7%	78.5%	83.8%
Naïve Bayes	77.9%	77.3%	80.8%

Fig. 5. Ten-fold cross-validation accuracy for bradycardia following Amiodarone administration.

Not surprisingly, Figure 4 recapitulates several well-known drug-drug interactions, such as Metoprolol and Pantoprazole, which are both metabolized by the liver enzyme CYP2D6[10]. Other entries represent obvious physiologic interactions, such as Amiodarone and Metoprolol, each of which acts independently to slow cardiac conduction (and thereby heart rate), while some drugs such as Aspirin likely function as a proxy for ICU patient stability. Interestingly, the importance of gender as a feature correlated roughly with the drug ranks from Figure 3.

Using the resulting feature sets, we evaluated each model using ten-fold cross-validation for all four drug-response combinations (Figures 5-8). Although none of our classifiers achieved an accuracy of greater than 90% for any single drug, each optimal model did considerably improve upon the baseline incidence rates (Amiodarone [32%], Metoprolol [34%, 31%], Furosemide [30%]), corresponding to the trivial KNN (k = n) classification. These results are consistent with those typically accepted in the context of the information loss inherent to the binarization of complex, continuous clinical outcome variables[11].

It is also noteworthy that in nearly all cases, the addition of non-pharmaceutical features improved the predictive ability of our models. This was most pronounced for Amiodarone and least so for Furosemide, consistent again with the relative strengths of their demographic-based differential responses. Surprisingly, in several cases, the non-pharmaceutical feature sets actually outperformed their training counterparts. These results in particular strongly support the argument for an expansion of drug-based decision support systems to incorporate readily-available, ancillary clinical information.

Metoprolol (bradycardia)			
	Pharmaceutical only	Non-pharmaceutical	All features
Logistic Regression	74.2%	76.4%	76.3%
Random Forest	79.5%	76.4%	79.6%
SVM	75.2%	76.4%	77.8%
Naïve Bayes	77.9%	77.3%	**80.8%**

Fig. 6. Ten-fold cross-validation accuracy for bradycardia following Metoprolol administration.

Metoprolol (hypotension)			
	Pharmaceutical only	Non-pharmaceutical	All features
Logistic Regression	74.1%	72.0%	74.3%
Random Forest	80.1%	72.0%	**80.2%**
SVM	75.5%	72.0%	76.2%
Naïve Bayes	77.4%	74.2%	77.2%

Fig. 7. Ten-fold cross-validation accuracy for hypotension following Metoprolol administration.

Furosemide (hypotension)			
	Pharmaceutical only	Non-pharmaceutical	All features
Logistic Regression	65.3%	66.7%	67.0%
Random Forest	73.3%	66.7%	**73.5%**
SVM	70.9%	66.7%	70.5%
Naïve Bayes	70.6%	67.3%	70.7%

Fig. 8. Ten-fold cross-validation accuracy for hypotension following Furosemide administration.

CONCLUSION

We have shown that it is possible to integrate large volumes of publically available human transcriptional data with open-access pharmacologic databanks to identify meaningful demographic-gene and demographic-gene-(protein)-drug associations. Further, we have demonstrated that the strength of these associations can be evaluated through the application of machine learning algorithms to a database of electronic health records, and that for specific candidate drugs, models incorporating these relationships outperform those using only the default input parameters (in this case, co-administered drug records).

Moving forward, we anticipate that the continued rapid expansion of the bioinformatics resources from which our method draws will serve to make such an approach increasingly valuable in the coming years.

ACKNOWLEDGEMENTS

We would like to thank our erstwhile CS 229 colleague Bob Arrigo for his sustained efforts throughout this project, as well as Professor Nigam Shah for facilitating our access to the MIMIC II database.

REFERENCES

1. Kemper, E.M., *et al.* Towards evidence-based pharmacotherapy in children. *Paediatr Anaesth* **21**, 183-189 (2011).
2. Parekh, A., Fadiran, E.O., Uhl, K. & Throckmorton, D.C. Adverse effects in women: implications for drug development and regulatory policies. *Expert Rev Clin Pharmacol* **4**, 453-466 (2011).
3. Mulla, H. Understanding developmental pharmacodynamics: importance for drug development and clinical practice. *Paediatr Drugs* **12**, 223-233 (2010).
4. Galanter, W.L., Didomenico, R.J. & Polikaitis, A. A trial of automated decision support alerts for contraindicated medications using computerized physician order entry. *J Am Med Inform Assoc* **12**, 269-274 (2005).
5. Xu, H., *et al.* Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. *J Am Med Inform Assoc* **18**, 387-391 (2011).
6. Overby, C.L., Tarczy-Hornoch, P., Hoath, J.I., Kalet, I.J. & Veenstra, D.L. Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. *BMC Bioinformatics* **11 Suppl 9**, S10 (2010).
7. Wheeler, D.L., *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **29**, 11-16 (2001).
8. Kohavi R, J.G. Wrappers for feature subset selection. *Artificial Intelligence* **1997**, 273-324 (1997).
9. Tamblyn, R., *et al.* A randomized trial of the effectiveness of on-demand versus computer-triggered drug decision support in primary care. *J Am Med Inform Assoc* **15**, 430-438 (2008).
10. Nozawa, T., *et al.* Influence of CYP2D6 genotype on metoprolol plasma concentration and beta-adrenergic inhibition during long-term treatment: a comparison with bisoprolol. *J Cardiovasc Pharmacol* **46**, 713-720 (2005).
11. Lin, F.P., Anthony, S., Polasek, T.M., Tsafnat, G. & Doogue, M.P. BICEPP: an example-based statistical text mining method for predicting the binary characteristics of drugs. *BMC Bioinformatics* **12**, 112 (2011).