의학석사 학위논문

MedCassandra : Personalized drug and ADR ranking forecast system based on personal genome variations

MedCassandra : 개인 맞춤 의학을 위한 유전체 기반의 약물 및 부작용 순위 예측 시스템

2013 년 2 월

서울대학교 대학원

의과학과

백 수 연

A thesis of the Master’s degree

MedCassandra : 개인 맞춤 의학을 위한 유전체 기반의 약물 및 부작용 순위 예측 시스템

MedCassandra : Personalized drug and ADR ranking forecast system based on personal genome variations

February 2013

The Department of Biomedical Sciences,
Seoul National University
College of Medicine
Su Youn Baik

ABSTRACT

Introduction: As the advancement of genome sequencing technology, there are some trials for pharmacogenomics interpretation. But traditional interpretation based on relationships of variant-drug response is simply annotating and listing variations and associated pharmacogenomic traits, not estimation and summarization of the total risk for each individual. So I develop MedCassandra, a system that forecast drug and ADR rank based on personal genome variations.

Methods & Results: MedCassandra consist of two parts. First I develop the algorithm for calculation of drug and ADR ranks. Second, I integrate drug information which requires for rank calculation from multi drug databases. When individuals input their genome variations, MedCassandra recommends the rank of drugs and ADRs that need to be cautious. As result of Asian individual's drug and ADR ranking, I confirm individual-specific drugs that individual should take more carefully. And overall individual's drug rank is reasonable when I compare highly ranked drugs to existing pharmacogenomics knowledge.

Conclusions: As MedCassandra recommend the alert list of drugs and ADRs, it can help that individual choose right drug

which they can use more safely and prepare and manage the ADR outbreak previously.

Keywords: Drug, Drug response, Adverse drug reaction, Genome variation, Personalized medicine

Student number: 2011-21941

CONTENTS

LIST OF TABLES AND FIGURES

1. INTRODUCTION

As the advancement of genome sequencing technology, the cost of genome sequencing has rapidly decreased. As the result, it is possible for individuals to sequence and analysis their genome. And issue of 'how to interpret personal genome' becomes important. There are many trials to interpret personal genome in clinical context. Ashley et al. undertook an integrated analysis of personal genome in variety point of view, like genetic risk prediction of disease, pharmacogenomics analysis and gene-environment interaction analysis (1). Karczewski et al. developed a system for private genome interpretation, Interpretome (2). When a user load a file which contains their genome variations to Interpretome's web page, Interpretome calculate disease risk or infer ancestry according to built-in analysis module.

Pharmacogenomics analysis is one of the most important topics in personal genome interpretation. This is because individual's genomic variations can influence drug response (3). Although people take same drug at same dose, some individuals obtain the desired effects but other individuals can have failure to drug efficacy or adverse drug reaction (ADR), even more death. So the goal of pharmacogenomics analysis is to realize

personalized medicine through prescribe 'right drug' to 'right patient' based on their genomic variations (4).

There are many studies for pharmacogenomics analysis. Many clinical studies based on case-control setting reveal associations between genomic variations and pharmacogenomic traits, like drug efficacy or ADR (5, 6). And pharmacogenomics Knowledge Base (PharmGKB, http://www.pharmgkb.org/) captures these relationships from literatures and ongoing pharmacogenetic researches (7). In PharmGKB, there are nearly 300 genomic variations which related to about 560 drug-response phenotypes. But pharmacogenomics interpretation based on relationships of variant-drug response is simply annotating and listing variations and associated pharmacogenomic traits, not estimation and summarization of the total risk for each individual (8).

In this study, I develop a system that forecast drug and ADR rank based on personal genomic variations named MedCassandra. The purpose of drug and ADR ranking is to make individual know which drug should be used with more caution and which ADR is changed frequency of occurrence because of their genomic variations. When individuals input VCF (Variant Call Format) file which contains their genomic variations in MedCassandra, drug and ADR rank is calculated.

So individual should be more careful when they using the drug which is highly ranked in their drug rank and the same for the ADR rank (Figure 1-1). This is help for people to select more safe drug which suitable to themselves.

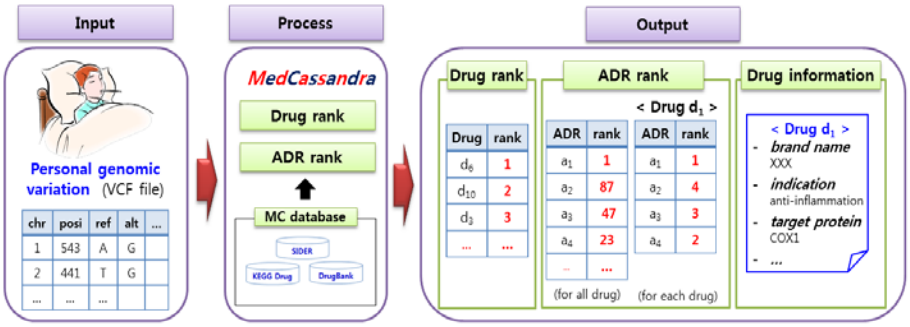

Figure 1-1. Overview of MedCassandra.

2. MATERIALS AND METHODS

2.1. Drug information integration

For ranking drug and ADR in MedCassandra, I integrate drug information like drug-gene relationship and drug-ADR relationship scattered around multiple drug-related databases. Process of drug information integration is parted in two steps. First step is that collecting drug entities registered in DrugBank (http://www.drugbank.ca/) (9) and KEGG Drug (http://www.genome.jp/kegg/drug/) (10) and integrating them

into MedCassandra(MC) drug entities. And second step is extracting and integrating drug information that need for drug and ADR ranking from DrugBank, KEGG Drug and SIDER (http://sideeffects.embl.de/) (11).

2.1.1. Integration of drug entity

There are different drug entities which are the same drug but are registered in different databases or redundantly registered in one database. For integrated these drug entities into one drug entity, I also extract drug names, synonyms, and chemical structures of each drug entities. Since different drug entities which have same active compound but different salt form are regarded as the same drug, I remove salt form from names, synonyms, and chemical structures of the drug entities on the basis of salt form list which extracted previous papers (12, 13). After that drug entities which satisfied any one of two requirements are regarded as same drug and integrate them into one MC drug entity.

REQUIREMENT 1. Chemical structures of drug entities are the same.

REQUIREMENT 2. Drug names or synonym of drug entities are same and they are counted as the same drug on the basis of manual curation

of pharmacist.

To get ADR information of integrated MC drug, I extract drug entities from SIDER which is the same in the MC drugs. To judge whether drug entities in SIDER is the same things in the MedCassandra, I use only drug names removed salt form, because there are no chemical structure of drug entities in SIDER.

After the integration process, I select drugs which are included in at least one criterion as followed,

(i) Drugs which are included in top 15 frequently prescribed drug classes during 2005~2008 in the United State (Health, United States, 2011, Centers for Disease Control and Prevention) (Table 2-1)

(ii) Drugs with information on pharmacogenomic biomarkers like gene variants or functional alterations which may alter drug response in their label (http://www.fda.gov/)

(iii) Drugs which have withdrawn from market (http://www.drugbank.ca/).

Table 2-1. Top 15 frequently prescribed drug classes during 2005~2008 in United State.

Drug Class	ATC code
Antihyperlipidemic agents	C10
Analgesics	N02
Antidepressants	N06A
Beta-adrenergic blocking agents	C07
Proton pump inhibitors	A02BC
ACE inhibitors	C09A
Sex hormones (contraceptives, menopause, hot flashes)	G03
Diuretics	C03
Thyroid drugs	H03
Antidiabetic agents	A10
Bronchodilators	R03
Anxiolytics, sedatives, and hypnotics	N05B, N05C
Antihypertensive combinations	C02
Calcium channel blocking agents	C08
Antihistamines	R06

2.1.2. Extraction of drug information

I extracted drug-gene relationship from DrugBank, KEGG Drug, and drug-ADR relationship from SIDER. There is no need to integrate MC drug's ADR information because it was described with standardized international medical terminology,

Medical Dictionary for Regulatory Activities (MedDRA). But a case of gene interaction information needs additional integration process. So first I extract gene symbol and KEGG human gene ID of genes which are target, transporter, enzyme and carrier of a drug from DrugBank and KEGG drug. After converting KEGG human gene IDs to UniProt ID using KEGG API (http://www.kegg.jp/kegg/docs/keggapi.html), I filtered out the genes which are not in human and unified gene names and symbols of human genes on the basis on human gene nomenclature provided by HGNC.

2.2. Drug and ADR ranking

2.2.1. Assumption of drug and ADR ranking

Ranking drug and ADR is based on the idea that if a gene (Gene A) is damaged because of personal genome variations (variants a and b), response of drugs (Drug 1, 2 and 3) which interacted with Gene A may be changed. And occurrence of ADR α also changed because ADR α is associated with Gene A through the three drugs (Drug 1, 2 and 3) which interact with the Gene A and at the same time have ADR α. So individuals who have genome variants a and b should be careful whether drug response or ADR α appear differently, when they take Drug 1, 2, and 3 (Figure 2-1).

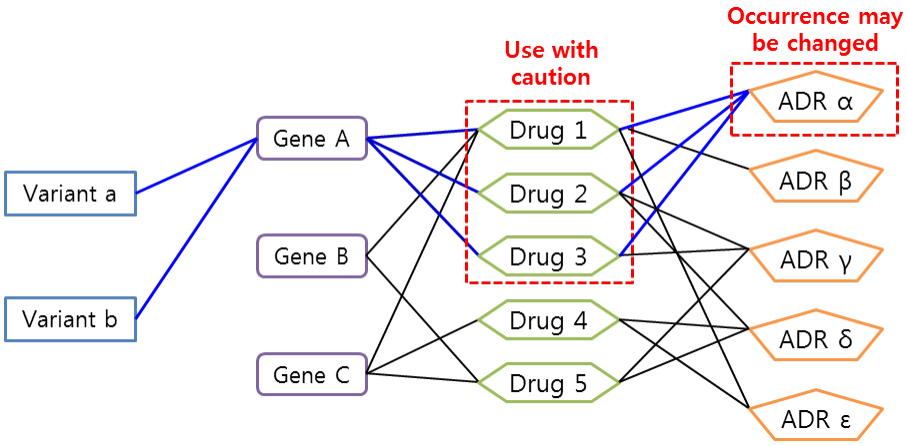

Figure 2 −1. Assumption of drug and ADR ranking

2.2.2. Drug and ADR ranking algorithm

Calculation of drug and ADR rank proceed through following two steps.

When individuals input their VCF file, I firstly evaluate the effects of the personal genome variations in VCF file on function of gene. Amino acid substitutions caused by non−synonymous variants which individual has can effect function of the gene and consequently influence response of drugs that interacted with the gene (14). And if more than two variants are in one gene, they synergistically work on the gene (15). For a given gene j which interacted a drug, I calculate a personal damaged gene score, s_{g_j}, according to formula as follows.

$$s_{g_j} = \left(\prod_{v_i \in G_j} \frac{\text{SIFT score}(v_i)}{\frac{1}{n}\sum_{p=1}^{n} \text{SIFT score}(v_{ip})} \right)^{1/|G_j|}$$

Here G_j is the set of variants which individual have and at the same time included in region of gene j. And $\text{SIFT score}(v_i)$ is the score that estimate the effects of a variant v_i on gene function using SIFT algorithm (14). I download pre-calculated SIFT scores of variants in human exome from the SIFT homepage (http://sift.jcvi.org/) and use it when I calculate s_{g_j}. Since the variants which are common in population are less likely to effect the function of gene (16), I adjust SIFT score by dividing SIFT score of a variant v_i by weighted average SIFT score of a variants at the same position for v_i in same ethnic group for individual. For calculation of weighted average SIFT score in each four major ethnic group, Asian, African, American and European, I used 1092 individual's VCF files provided from 1000Genome project (http://www.1000genomes.org/)(17). SIFT score of a variant is lower when their effects on function of gene are more significant, so damaged gene score is lower when gene is more damaged.

Then I calculate personal drug score which measure the degree of probability of drug-response change. If genes which related to drug action are functionally altered because of genomic variations, they synergistically affect drug response.

So for given drug k, drug score, s_{d_k}, is calculated according to formula as follows.

$$s_{d_k} = \left(\prod_{g_j \in D_k} s_{g_j} \right)^{1/|D_k|}$$

Here, D_k is the set of genes which interacted with drug k. As damaged gene score is lower when gene is more damaged, drug score is lower when response of the drug is more likely to be changed. Then I rank drugs in increasing order of drug score. Highly ranked drug means individual can experience change of drug response more frequently when they using the drug.

Like drug score, for ADR l, personal ADR score, s_{a_l}, is calculated according to formula as follows.

$$s_{a_l} = \left(\prod_{g_j \in A_l} s_{g_j} \times w_{jl} \right)^{1/|A_l|}$$

Here, A_l is the set of genes which associated with ADR i. And w_{jl}, a weight which represent a degree of association between gene j and ADR l, is calculated the fraction of drugs which interacted with gene j and have ADR l at the same time. Then I rank ADRs in increasing order of ADR score. Highly ranked ADR means individual should be more careful about outbreak of the ADR.

2.3. Validation of drug ranking

To validate individual's drug and ADR rank, I compare the individual's drug and ADR rank with the existing pharmacogenomics knowledge. So I collected the variant-drug associations from PharmGKB.

After comparing variants in individual's genome with the variants list from the existing pharmacogenomics knowledge, I extract list of drugs which influenced by the variants that individual have. The drug list adopted as a gold standard, I calculated precision, recall of individual's drug throughout all rank thresholds to measure gold standard is ranked highly in results of MedCassandra.

3. RESULTS

3.1. Drug information integration

6,708 drug entities are extracted from DrugBank and 9,777 drug entities from KEGG drug. Total 16,485 drug entities are integrated into 12,770 MC drug entities. And 963 drug entities which 94.0% of 996 drug entities in SIDER are mapped to 986 MC drug entities. Also after extracting 2,329 human genes from DrugBank and 527 human genes from KEGG drug, total 2,435

human genes which interact with 12,770 MC drug entities are listed.

The number of drug entities from DrugBank or KEGG drug which are integrated one MC drug entity vary from 1 to 17 (Figure 3-1). This tells us the necessity of drug entity integration at the next 2 points. First 82.2% of MC drug entity is derived from only one drug entity of DrugBank or KEGG drug. It means different drug is registered in each two major drug databases. And 5.7% of MC drug entity is derived from 2 or more drug entities of one drug database. It means there are redundant drug entities registered in each databases.

Among 12,770 MC drug entities, only 548 drug entities are meet the criteria mentioned above section. Then I remove 47 drugs that do not have any interacted gene. As the result, 501 drugs are selected and used in drug and ADR ranking. Integration statistics of MedCassandra is summarized Table 3-1.

	(1)	(2)
The number of drug	12,770	501
The number of ADR term	4,434	3,284
The number of human gene interacted with drugs	2,435	551
The number of drug-gene relationship	18,031	4,215
The number of drug-ADR relationship	119,509	47,547

Table 3-1. Integration statistics of MedCassandra.

(1) Total statistics of the drugs and their information integrated in MedCassandra. (2) Among MC drugs, statistics of the drugs and their information that used in drug and ADR ranking.

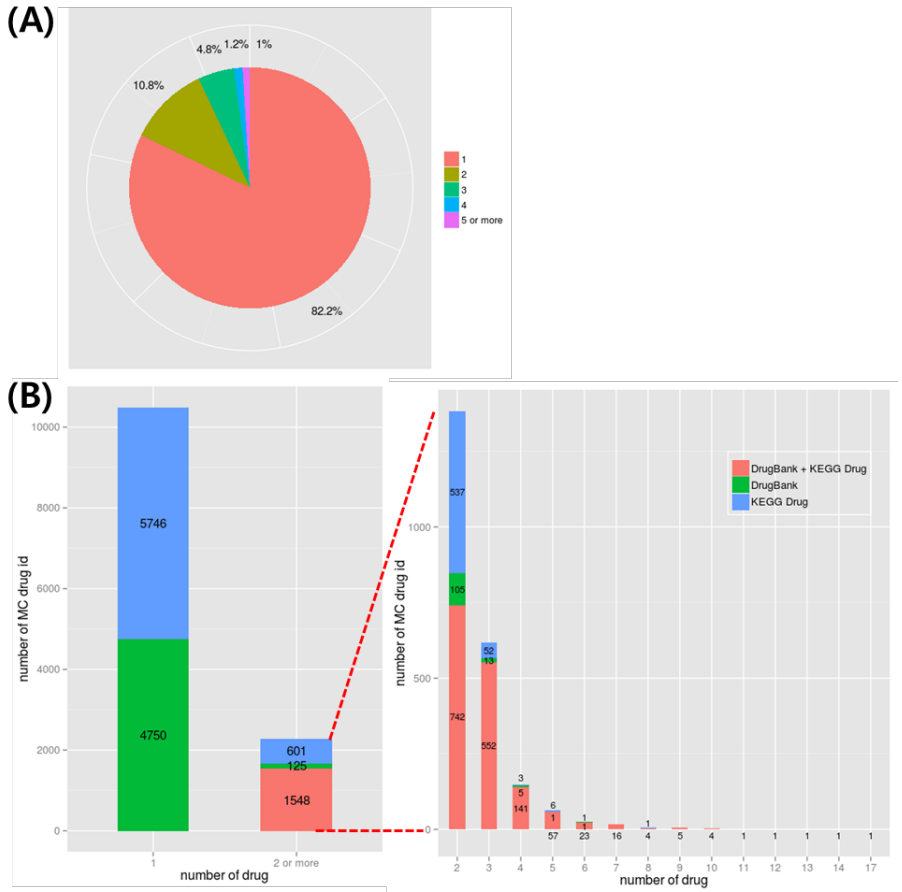

Figure 3-1. Statistics of drug entity integration.

(A) Distribution of the number of MC drug entity according to the number of drug entity included in one MC drug entity. (B) Bar type of graph (A). x axis means the number of drug entity included in one MC drug entity. Blue bar is the number of MC drug entity only from the KEGG drug entities, green bar is only from the DrugBank drug entities. And red one is from both databases.

3.2. Result of drug and ADR ranking for individuals

I applied drug and ADR ranking algorithms to one Korean individual, called individual A. There are 3,309,258 variants in A's genome. Among them, 670 (0.02%) variants are in the exon region of 313 (56.8%) drug-related genes. As the result of the drug ranking, top 15 ranked drugs are listed in Table 3-2. According to result, top ranked drug is Meclizine, anti-histamine agent. And individual A needs caution when using anti-diabetics drugs, like insulin glulisine, insulin aspart, saxagliptin. When she takes these drugs, she may have a chance to experience change of drug response or occurrence of ADR.

But the drug rank is more meaningful when compare the drug in same category. So I calculated rank of 15 drugs in lipid modifying agents. Individual A's drug rank is listed in Table 3-3. The drug probucol is top ranked in individual A. And when A takes lipid modifying agents, it is to recommend to choice niacin or fibrate except for fenofibrate, rather than HMG CoA reductase inhibitors.

Rank	Drug name	Drug score
1	Meclizine	1.00E-04
1	Ximelagatran	1.00E-04
1	Insulin Glulisine	1.00E-04
1	Insulin Aspart	1.00E-04
1	Omalizumab	1.00E-04
6	Nitroprusside	0.000969
7	Beclomethasone	0.001658
8	Temafloxacin	0.001778
9	Frovatriptan	0.002154
9	Saxagliptin	0.002154
9	Insulin Detemir	0.002154
9	Insulin Glargine	0.002154
9	Insulin Lispro	0.002154
14	Betaxolol	0.006225
15	Hydroxyzine	0.00625

Table 3-2. Top 15 drug rank of individual A.

Rank	Drug name	Drug score
1	Probucol	0.03208
2	Rosuvastatin	0.068925
3	Gemfibrozil	0.071572
4	Fluvastatin	0.080292
5	Fenofibrate	0.156683
6	Simvastatin	0.163921
7	Lovastatin	0.181083
8	Cerivastatin	0.182276
9	Pravastatin	0.237502
10	Atorvastatin	0.246433
11	Levothyroxine	0.614117
12	Ezetimibe	0.635155
13	Clofibrate	0.679218
14	Bezafibrate	0.683902
15	Niacin	0.706219

Table 3-3. Drug rank in lipid modifying agent of individual A.

As the result of the ADR ranking, individual A's top 10 ranked ADRs are listed in Table 3-4. As same for drug rank, A's top ranked ADRs are associated with glucose metabolism, like diabetes mellitus, insulin allergy and background diabetic retinopathy. So when she takes drug, it is recommended for her to pay attention whether these symptoms occur or not.

Rank	ADR name	ADR score
1	type 1 diabetes mellitus	9.97E-05
2	burn local	0.001654
2	thermal burn	0.001654
2	oropharyngeal candidiasis	0.001654
5	background diabetic retinopathy	0.002149
5	insulin allergy	0.002149
7	salmonella sepsis	0.00215
8	type 2 diabetes mellitus	0.004541
9	breast abscess	0.006212
9	punctate keratitis	0.006212

Table 3-4. Top 10 ADR rank of individual A.

Rank	Drug name	ATC code (level 1)
1	Iloperidone	N
2	Remoxipride	N
3	Fexofenadine	R
4	Tegaserod	A
5	Cyproheptadine	R
6	Cisplatin	L
7	Methysergide	N
8	Cisapride	A
9	Epinastine	RS
10	Omeprazole	A
11	Fenfluramine	A
12	Sertindole	N
13	Minaprine	N
14	Nedocromil	RS
15	Fluoxetine	N

Table 3-5. Top 15 drug rank of Asian.

We also calculate drug rank of 286 Asian population using 1000Genome Asian's VCF files. After calculate drug rank in each person, I aggregate drug ranks into one drug rank with Borda count rank aggregation method (18). In top 15 ranked drug, there are many drugs acting on nervous system such as iloperidone, remoxipride and methysergide (Table 3-5). This tendency is sustained when I count drug frequency in each therapeutic class at each rank interval (Figure 3-2). In top 1~100 ranked drugs, drugs acting on nervous system occupy large proportion than other drug classes and the proportion is decreased as rank is lower. Conversely proportion of drugs acting on respiratory system is higher as drug rank is lower.

I validate drug ranking algorithm in Asian population using existing pharmacogenomics knowledge. For each 286 Asian individual in 1000Genome, I calculate precision and recall throughout all rank thresholds. After average them at each threshold, I draw receiver operating characteristic (ROC) curve and calculate area under the curve (AUC) (Figure 3-3). Average of AUC is 0.634 (±0.002) for drug rank. This result is higher than 0.5, so my algorithm is reasonable to predict personal drug responsibility.

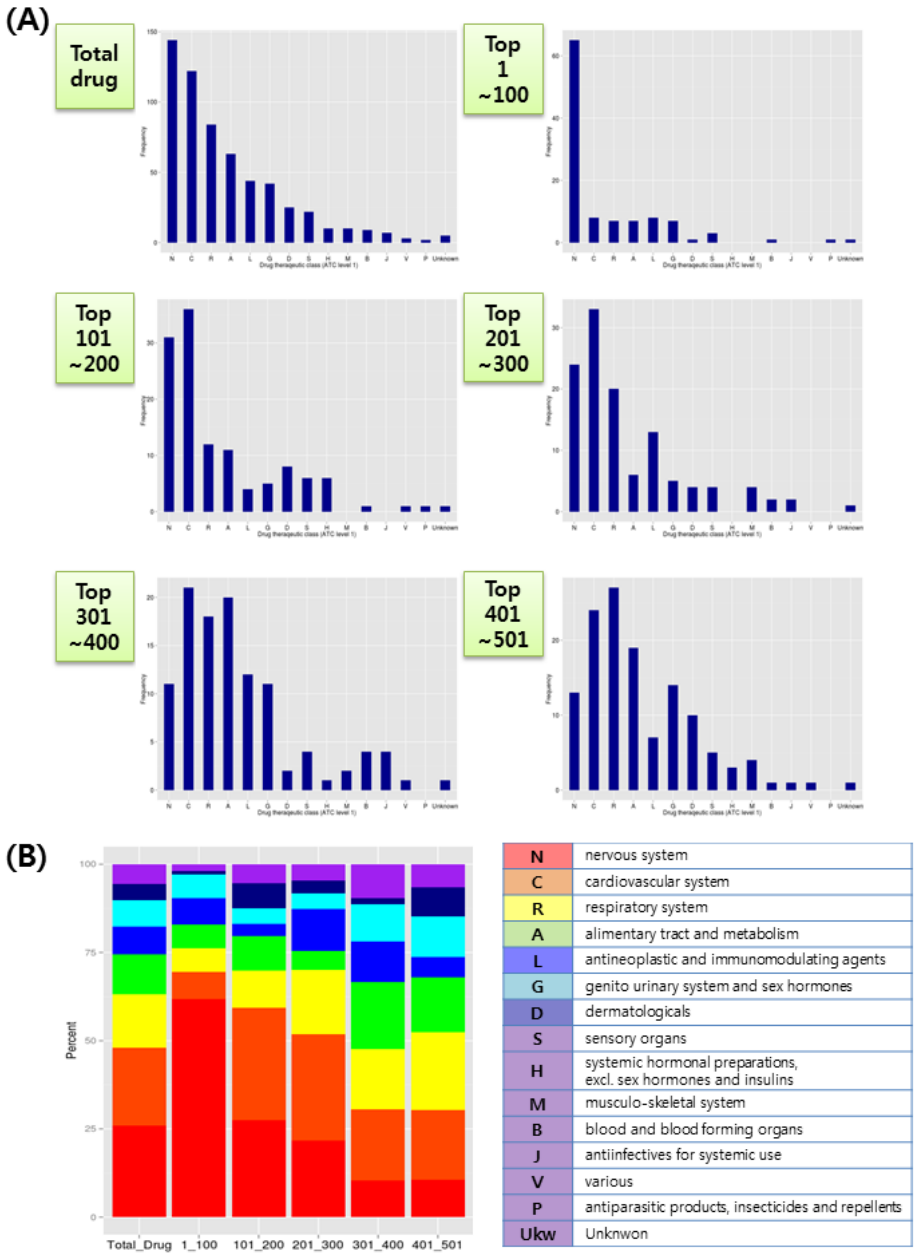

Figure 3-2. Drug frequency in each therapeutic class at each rank interval

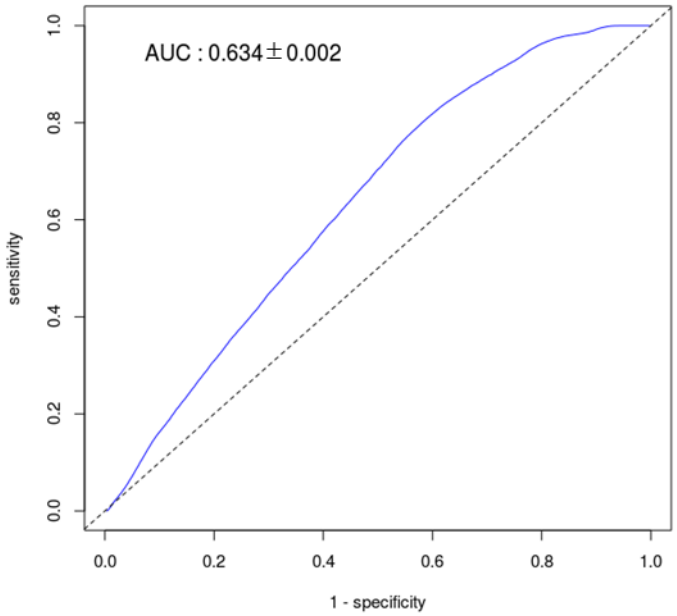

Figure 3-3. ROC curve for drug rank validation in Asian population.

4. DISCUSSION

As using MedCassandra, individuals can know which drug is used more carefully based on their genomic variation. It can be help people to choose the drug that they can use more safely. And as individuals previously know the ADRs that more frequently occurred, they can prepare and manage the ADR outbreak more easily. These points are the major contribution of MedCassandra to personalized medicine. And these whole processes are needed just the existing pharmacological

knowledge, drug-gene interaction information and drug-ADR relationship, without large scaled clinical study like GWAS.

But MedCassandra also have some limitations. First, result of drug and ADR ranking depends on drug-gene and drug-ADR information. Rank is not calculated in case of drugs that their interaction partner is not discovered yet. And if new drug-gene interaction is discovered or ADR is newly outbreak, results of drug and ADR rank will be changed. Second, some genomic variants change drug response positively, and others negatively. For example of clozapine, people who have rs742105(C>T) variant in DTNBP1 gene have better response to clozapine (19). However people who have variation in CYP1A2 gene (CYP1A2*1F) are poor responder (20). But MedCassandra cannot discriminated these opposite effect, just say, 'response of clozapine may be changed. Finally, ranking in MedCassandra is only based on the personal genomic profile. But non-genetic factor including age, height, race is also the important factor when estimate response of some drugs like warfarin (21). MedCassandra does not cover the effects of non-genetic factor on drug response.

REFERENCES

1. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, et al. Clinical assessment incorporating a personal genome. Lancet. 2010;375(9725):1525-35. Epub 2010/05/04.

2. Karczewski KJ, Tirrell RP, Cordero P, Tatonetti NP, Dudley JT, Salari K, et al. Interpretome: a freely available, modular, and secure personal genome interpretation engine. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2012:339-50. Epub 2011/12/17.

3. Johnson JA. Pharmacogenetics: potential for individualized drug therapy through genetics. Trends in genetics : TIG. 2003;19(11):660-6. Epub 2003/10/31.

4. Pagliarulo V, Datar RH, Cote RJ. Role of genetic and expression profiling in pharmacogenomics: the changing face of patient management. Current issues in molecular biology. 2002;4(4):101-10. Epub 2002/11/16.

5. Thervet E, Loriot MA, Barbier S, Buchler M, Ficheux M, Choukroun G, et al. Optimization of initial tacrolimus dose using pharmacogenetic testing. Clinical pharmacology and therapeutics. 2010;87(6):721-6. Epub 2010/04/16.

6. Bouman HJ, Schomig E, van Werkum JW, Velder J, Hackeng CM, Hirschhauser C, et al. Paraoxonase-1 is a major

determinant of clopidogrel efficacy. Nature medicine. 2011;17(1):110-6. Epub 2010/12/21.

7. Hernandez-Boussard T, Whirl-Carrillo M, Hebert JM, Gong L, Owen R, Gong M, et al. The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge. Nucleic acids research. 2008;36(Database issue):D913-8. Epub 2007/11/23.

8. Cordero P, Ashley EA. Whole-Genome Sequencing in Personalized Therapeutics. Clinical pharmacology and therapeutics. 2012. Epub 2012/05/03.

9. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic acids research. 2011;39(Database issue):D1035-41. Epub 2010/11/10.

10. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic acids research. 2006;34(Database issue):D354-7. Epub 2005/12/31.

11. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Molecular systems biology. 2010;6:343. Epub 2010/01/21.

12. Paulekuhn GS, Dressman JB, Saal C. Trends in active pharmaceutical ingredient salt selection based on analysis of the Orange Book database. Journal of medicinal chemistry. 2007;50(26):6665-72. Epub 2007/12/07.

13. Serajuddin AT. Salt formation to improve drug solubility. Advanced drug delivery reviews. 2007;59(7):603-16. Epub 2007/07/10.

14. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annual review of genomics and human genetics. 2006;7:61-80. Epub 2006/07/11.

15. Jelier R, Semple JI, Garcia-Verdugo R, Lehner B. Predicting phenotypic variation in yeast from individual genome sequences. Nature genetics. 2011;43(12):1270-4. Epub 2011/11/15.

16. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012;335(6070):823-8. Epub 2012/02/22.

17. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73. Epub 2010/10/29.

18. Dwork C, Kumar R, Naor M, Sivakumar D, editors. Rank aggregation methods for the web. Proceedings of the 10th international conference on World Wide Web; 2001: ACM.

19. Zuo L, Luo X, Krystal JH, Cramer J, Charney DS, Gelernter J. The efficacies of clozapine and haloperidol in refractory schizophrenia are related to DTNBP1 variation.

Pharmacogenetics and genomics. 2009;19(6):437-46. Epub 2009/04/17.

20. Ozdemir V, Kalow W, Okey AB, Lam MS, Albers LJ, Reist C, et al. Treatment-resistance to clozapine in association with ultrarapid CYP1A2 activity and the C-->A polymorphism in intron 1 of the CYP1A2 gene: effect of grapefruit juice and low-dose fluvoxamine. Journal of clinical psychopharmacology. 2001;21(6):603-7. Epub 2002/01/05.

21. Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. The New England journal of medicine. 2009;360(8):753-64. Epub 2009/02/21.

국문 초록

서론: 유전체 서열 분석 기술이 발전함에 따라, 개인이 가진 유전체 변이를 바탕으로 개인의 약물 반응성 변화를 예측하고 개인에게 약물 유전학적 해석을 제공하려는 연구들이 진행되고 있다. 하지만 유전체 변이와 약물 반응 간의 관련성 정보를 기반으로 하는 기존 약물 유전학적 해석은 단순히 개인이 가진 각각의 유전자 변이에 따른 약물 반응성의 변화를 나열해 줄 뿐, 개인의 약물 반응성을 종합적으로 예측 및 해석해 줄 수 없다. 본 연구에서는 MedCassandra 라고 하는 개인의 유전체 변이를 기반으로 한 약물 및 부작용 순위 예측 시스템을 계발하였다.

방법 및 결과: MedCassandra 는 크게 개인이 입력한 유전체 변이에 따라 주의해야 할 약물과 부작용을 순위화하는 알고리즘과 이를 위한 약물 지식의 통합이라는 두 가지 측면으로 구성된다. 개인이 자신의 유전체 변이를 입력하면, MedCassandra 는 개인이 주의해야 할 약물과 부작용 목록을 순위화하여 제공한다. 1000 Genome project 에서 제공하는 아시아 인구 집단의 유전체 변이를 입력 값으로 하여, 개인에게 위험한 약물을 순위화해 본 결과, 개인별로 주의해야 할 약물들이 차별화 되는 것을 확인할 수 있었으며, 이는 기존의 약물 유전학적 결과와도 어느 정도 유사한 것으로 나타났다.

결론: MedCassandra 은 개인이 주의해야 할 약물과 부작용 목록을 미리 제공해 줌으로써, 개인이 자신에게 맞는 약물을 선택하고 발생 가능한 부작용에 대해서는 미연에 대비 및 대처할 수 있도록 도울 수 있다.

주요어 : 약물, 약물 반응, 약물 부작용, 유전체 변이, 개인 맞춤 의학

학 번 : 2011-21941