Pharmacogenetics of advanced colorectal cancer treatment

Pander, J.

7

Pharmacogenetics of tomorrow: the 1+1=3 principle

Jan Pander • Judith A.M. Wessels • Ron H.J. Mathijssen • Hans Gelderblom
Henk-Jan Guchelaar

Pharmacogenomics 2010 Jul;11(7):1011-7

Summary

Disappointing results from replicating pharmacogenetic association studies have prompted the search for novel statistical techniques to analyze the data, while taking into account the biological complexity underlying drug response. Two of these techniques – multifactor dimensionality reduction and classification and regression tree – will probably be applied in increasing numbers of future pharmacogenetic studies. In this article, we describe the concepts underlying both techniques and illustrate their application in a recent pharmacogenetic study.

Pharmacogenetic studies aim at predicting drug response. These studies commonly test associations between single candidate genetic polymorphisms and the efficacy or toxicity of drugs. Often, genetic polymorphisms in genes that have a putative impact on the function of the corresponding protein are selected. For their part, these proteins are assumed to have an impact on drug response, being enzymes involved in the pharmacokinetics or pharmacodynamic targets of the drug of interest. Each polymorphism is then separately associated with drug efficacy or toxicity.

Unfortunately, initial results from these candidate gene approach studies are often not replicated in subsequent studies.[1] This is clearly illustrated in large pharmacogenetic studies across different diseases.[2-4] Even when a study is successfully replicated, the effect of a polymorphism on drug response is often lower than initially described.[5] This is one of the reasons that only a handful of pharmacogenetic markers are actually useful to individualize treatment in clinical practice.[6]

An explanation for the disappointing results could be that the classic candidate gene approach does not take into account the full complexity underlying drug response. Drug response is likely to be influenced simultaneously by different biochemical components, such as pharmacokinetic enzymes and molecular targets within a biochemical pathway. Furthermore, it is recognized that the interplay between these different molecular components is extensive and complex. From a biological point of view, it seems therefore not only appropriate to study polymorphisms in candidate genes collectively – the so called candidate pathway approach[7] – but also to assess the interaction between the polymorphic genes. This interaction means that the impact attributed to one genetic polymorphism depends on one or more others.[8] In some cases, haplotype analysis can reveal relevant but simple interactions between polymorphisms, such as combined analysis of *CYP3A4* and *CYP3A5* variation for docetaxel pharmacokinetics.[9] However, for genes that are located on different chromosomes, haplotype analysis is usually not possible.

Genetic interaction studies have already been published investigating susceptibility to several complex diseases; thus, the concept itself not new.[10-15] However, the application of this concept in pharmacogenetic studies is scarce.[16-18]

Since results from the candidate gene approach have been disappointing [2-4], and because the biologic rationale supports studying gene-gene interactions, we anticipate that novel techniques for analysis will be applied to pharmacogenetic studies in the near future.

To determine which interactions are most important for drug response, statistical techniques must be used. The most widely used technique in genetics is (logistic) regression analysis with interaction. The advantage of this technique is its availability in common statistical packages, and that covariate adjustments can be made in the same analysis. However, assumptions on the genetic model must be made beforehand, which may not be accurate in complex interaction analysis. Moreover, (logistic)

regression analysis is of only limited application with increasing numbers of polymorphisms, as the number of possible interactions increases substantially with increasing numbers of polymorphisms. For instance, the total number of possible two-, three- and four-way interactions for ten polymorphisms is 375, whereas it is more than 4,000,000 for 100 polymorphisms. This illustrates the complexity of the interaction analysis, and has lead to the application and development of more advanced techniques for interaction analysis.

These techniques rely on algorithms that reduce the number of dimensions – that is, possible combinations of polymorphisms – in order to establish a genetic classifier to predict drug response. An important aspect of these genetic classifiers is that the combination of different polymorphisms results in information gain (the 1+1=3 principle). This concept of synergy illustrates the impact of interaction most intuitively, but it must be noted that other types of interaction exist, as reviewed by Perez-Perez et al.[19]

Two of these advanced techniques will be described: 'multifactor dimensionality reduction' (MDR) [20,101] and 'classification and regression tree' (CART) analysis, for their application in pharmacogenetics. Since these methods use different and unique approaches, we have no explicit preference. To illustrate the application of both techniques, genetic classifiers were created to predict the incidence of leukopenia (grade 0 *versus* grade ≥1 according to the National Cancer Institute Common Toxicity Criteria) in patients treated with single-agent sunitinib, using data from a candidate gene analysis for sunitinib induced toxicity.[21] In this cohort of 198 Dutch patients (predominantly with metastatic renal cell carcinoma and gastrointestinal stromal tumors) who were assessable for leukopenia, 31 polymorphisms were analyzed in 12 genes that encode enzymes in the pharmacokinetic and pharmacodynamic pathways of sunitinib. Genotyping was performed on the Biomark™ 48.48 Dynamic Array (Fluidigm, San Francisco, CA, USA) using Taqman® assays (Applied Biosystems, Nieuwekerk aan den IJssel, the Netherlands) according to the manufacturer's protocol as previously described.[21] There are more than 36,000 possible two-, three- and four-way interactions possible for these 31 polymorphisms, emphasizing the complexity of the problem and the need for advanced statistical techniques.

Multifactor dimensionality reduction

The theoretical application of the MDR analysis software to pharmacogenetic studies has previously been reviewed by Ritchie et al.[22] The essence of the MDR analysis is that all possible combinations of genotypes are evaluated to predict drug response. Each combination of genotypes contains a ratio of responders to non-responders, which is used to classify patients. In this way, the complexity of genotype combinations

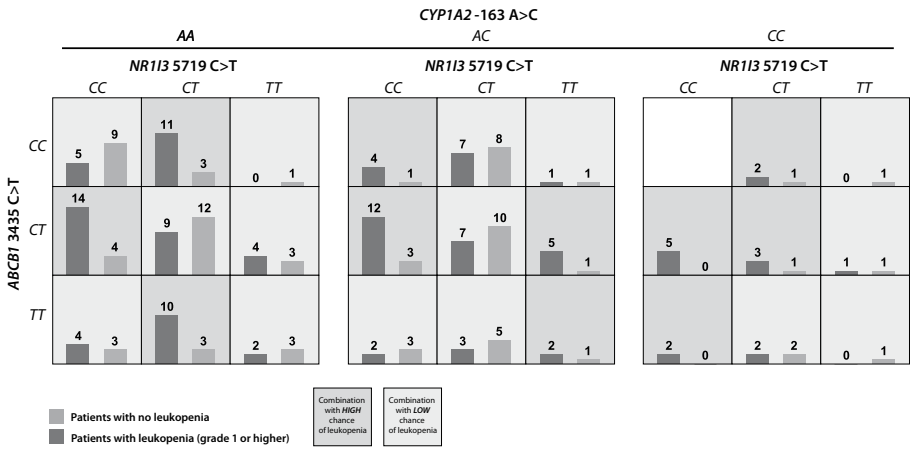
– dimensions – is reduced into a straightforward high-/low-risk factor. From all the possible combinations of genotypes, the MDR method presents the combination – or genetic classifier – that predicts drug response or toxicity the best.

The MDR method selects the best genetic classifier based upon accuracy – or lowest classification error. Genotypes that are individually associated with drug response contribute to a higher accuracy, and will therefore usually end up in the best genetic classifier, while it is uncertain whether they provide substantial information gain. However, it is possible that a polymorphism with a main effect also contributes substantially to the interaction model. Therefore, excluding polymorphisms could be disadvantageous. Currently, the MDR software is being updated, so that it is possible to adjust for main effects of individual polymorphisms. However, as the method of covariate adjustment has not proved its value, we excluded individually associated polymorphisms from the MDR analysis. Furthermore, haplotypes disturb the selection of the best genetic classifier because of over-fitting due to the increased number of genotype groups based upon haplotype combinations. Therefore, preferably only single nucleotide polymorphisms should be included in the analysis. In the example of our previous sunitinib analysis, polymorphisms in *NR1I3* (7738A>C and 7837T>G), *VEGFR* (-92G>A and 1718T>A), *CYP1A1* (2455A>G) and *FLT3* (738T>C) had to be excluded from the analysis because of their individual associations. Next, every possible combination of genotypes is evaluated, and the software computes how well the best genetic classifier predicts drug response. The analysis is performed across tenfold cross-validation samples to correct for over-fitting, and the combination with the highest accuracy in the cross-validation is considered the best genetic classifier. In our example, the combination containing three polymorphisms, *NR1I3* 5719C>T, *ABCB1* 3435C>T and *CYP1A2* -163A>C, showed the highest accuracy of 61.8% (*P*=.008 obtained by permutation [23,102]) which means that the average classification error in the prediction sets from cross-validation is 38.2%. Other combinations of polymorphisms resulted in lower accuracies. The distribution of patients with and without leukopenia across the three polymorphisms is shown in figure 1A. The interaction dendrogram for this genetic classifier is shown in figure 1B. The orange and red lines indicate a synergistic interaction between the polymorphisms. The short red lines between the *ABCB1* and *NR1I3* polymorphisms indicate that the interaction between these polymorphisms is the strongest in this model.

The results can be used to create a genetic classifier of response. This classifier can then be used in regular statistical analysis to compute an odds ratio (OR), and to perform a multivariate analysis. In the sunitinib example, the multivariate logistic regression analysis was performed including age, gender, WHO performance status, the genetic classifier, and the polymorphisms that were individually associated with leukopenia. The genetic classifier obtained by MDR has a corrected OR of 4.06 (95% confidence interval (CI), 1.99 to 8.31), whereas only the polymorphisms in *CYP1A1* and

Figure 1 Multifactor dimensionality reduction analysis of sunitinib induced leukopenia

A



B

(**A**) The genetic classifier consisting of polymorphisms in *CYP1A2*, *NR1I3* and *ABCB1* resulted in the highest accuracy of 61.8% in the cross-validation sample. For each genotype combination, the number of patients with and without leukopenia is shown. Combinations with low chance of leukopenia are shaded light grey, whereas combinations with high chance of leukopenia are shaded dark grey. Since in the total group of patients, 59.1% experienced leukopenia, a combination is considered to give a high chance of leukopenia when the percentage of patients experiencing leukopenia exceeds 59.1%. (**B**) Interaction dendrogram for the polymorphisms included in the genetic classifier obtained by multifactor dimensionality reduction. There was synergistic interaction, with the strongest interaction between the *ABCB1* and *NR1I3* polymorphisms.

FLT3 remained statistically significant (*P*=.043 and *P*=.010, respectively) in the multivariate analysis.

When a logistic regression analysis was performed with the three-way interaction between the *NR1I3* 5719C>T, *ABCB1* 3435C>T and *CYP1A2* -163A>C polymorphisms, the interaction term was not significantly associated with leukopenia. Each polymorphism was included as an ordinal factor, whereas the MDR method did not rely on this *a priori* assumption. This underlines the fundamental difference between these two methods, besides the fact that every possible interaction could not be assessed using logistic regression.

When the polymorphisms with a main effect were also included in the MDR analysis, all top models contained at least either the *NR1I3* 7738A>C or *FLT3* 738T>C polymorphism.

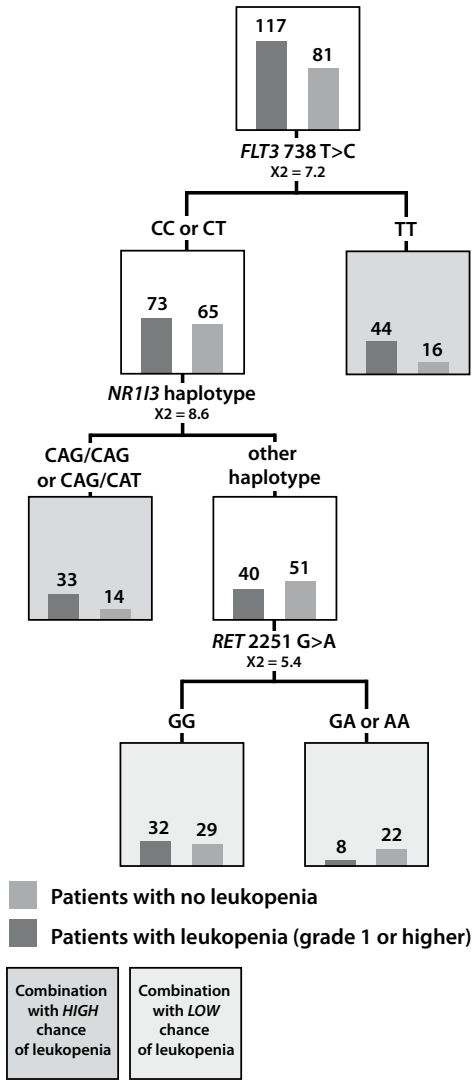
Classification and regression tree

The essence of the CART analysis is that patients are divided into groups with a unique genotype combination that predicts drug response. During the CART analysis, patients are subdivided successively, in such a way that a so-called 'classification tree' is grown. Each subgroup is split by the most discriminating polymorphism, which could be a different polymorphism for each subgroup. This procedure is continued until the pre-specified maximum tree depth is reached, or when each subgroup reaches a pre-specified minimum number of patients. Each terminal subgroup of the tree contains a ratio of responders to non-responders, which can be used to classify patients. Since each subgroup can be split by a different polymorphism, interaction can be detected, meaning that the influence of each polymorphism depends on the polymorphisms that split the subgroup in a previous level of the tree.

For the CART analysis, polymorphisms that are associated individually with drug response can be included, as well as haplotypes. In the sunitinib example, all polymorphisms and haplotypes were included as previously reported.[21] The maximum tree depth was set to three levels, and no subgroup was allowed to contain less than 25 patients. Each subgroup was split based upon the highest χ^2 value. In figure 2, the classification tree is shown. The tree contained the polymorphism in *FLT3* (738T>C, step 1), the haplotype in *NR1I3* (step 2) and the polymorphism in *RET* (2251G>A, step 3). Each terminal group of the tree can be seen as a unique combination of genotypes – a genetic classifier. As in the MDR analysis, the genetic classifier was used in a regular statistical analysis to compute the OR. In a logistic regression analysis including WHO performance status, age and gender, the OR for the genetic classifier was 3.36 (95%CI, 1.84 to 6.15).

When a logistic regression analysis was performed with the three-way interaction between the *FLT3* 738T>C and *RET* 2251G>A polymorphisms and the *NR1I3* haplotype, the interaction term was not significantly associated with leukopenia, again emphasizing the fundamental difference between these two methods.

Figure 2 Classification and regression tree analysis of sunitinib-induced leukopenia



Each branch of the tree is divided by the polymorphism or haplotype with the highest χ^2 value. Terminal groups are shaded light grey or dark grey for relatively low and high risk of sunitinib-induced leukopenia, respectively. Since 59.1% of the patients experienced leukopenia (grade 1 or higher), a genetic classifier is considered high chance of leukopenia when the percentage of patients experiencing leukopenia exceeds 59.1%.

Interpretation and validation of the genetic classifier

Both MDR and CART analyses result in genetic classifiers that are associated with drug efficacy or toxicity. Since these methods rely on different ways to create this genetic classifier, they result in different classifiers that do not necessarily contain the same polymorphisms. In the MDR analysis, a genetic classifier is created in addition to polymorphisms that were individually associated with drug response, so that the genetic contribution to drug response is further explored. In the CART analysis, the genetic contribution to drug response is analyzed taking into account that a polymorphism may only have impact on drug response under the condition that another polymorphism is present. The similarity between the methods is that combinations of genotypes are investigated, rather than individual polymorphisms. This is more plausible from a biological point of view, because drug response is a complex trait and involves many proteins. Importantly, the CART and MDR methods detect statistical interaction, and the models do not necessarily contain polymorphisms in genes encoding enzymes that interact biologically. The interpretation of the genetic classifiers from a biological point of view is therefore not straightforward. The genetic classifiers contain polymorphisms that only exert their influence under the condition that other polymorphisms are present. In the MDR analysis of our sunitinib example, three polymorphisms in metabolic enzymes were included in the genetic classifier. From a biological point, it is likely that metabolic routes compete, and that the effect of one polymorphism on the metabolic capacity can be altered by others. When interpreting the genetic classifier obtained by CART in our sunitinib example, it appears that genetic variation in the metabolic enzyme NR1I3 is only relevant for carriers of the *FLT3* 738C-allele and not for carriers of the *FLT3* 738TT genotype, possibly because the latter are more sensitive to sunitinib-induced leukopenia regardless of the plasma levels of sunitinib.

Critical choices have to be made before these techniques can be applied, such as the number and selection of patients, the selected polymorphisms and the settings of the software. Importantly, when large numbers of polymorphisms are included in the interaction analysis, the number of possible interactions becomes enormous. In the current era of whole-genome profiling of more than a million polymorphisms, intelligent filtering of polymorphisms must be performed before interaction analysis, due to the computational requirement of such analysis.[24] Furthermore, both MDR and CART may result in genetic classifiers that predict drug response in the original patients better than in new patients because of potential over-fitting. The ORs for the genetic classifiers in our examples are therefore likely to be biased, and the true OR has to be obtained in an independent validation cohort. For these genetic classifiers to be applied in clinical practice, the genetic classifier should therefore be confirmed in independent cohorts.[25] Before the effort of external validation is undertaken,

internal validation can be performed to correct for over-fitting using for instance cross-validation.

Conclusion

Statistical techniques to analyze high-order interactions between polymorphisms, such as the MDR and CART techniques, create genetic classifiers that predict drug response. They have the major advantage over classic pharmacogenetic association studies that the complexity underlying drug response is studied and may therefore be more likely to be successfully replicated. When validated, these genetic classifiers can provide a novel and more rational approach to individualizing drug treatment.

Future perspective

We believe that complex interaction between polymorphisms will increasingly be studied in the near future, since the results from traditional pharmacogenetic association studies have been disappointing. The MDR and CART methods will probably be the most widely used, as they are widely available and relatively easy to apply. However, for the resulting genetic classifiers to reach the clinic, thorough validation must be performed using independent patient populations. Only when validation has been successful can the genetic classifiers be used to guide individualized therapy.

Executive summary

Background

- Recent pharmacogenetic association studies on frequently studied polymorphisms failed to replicate initial findings.
- Drug response is a complex phenomenon, and involves many different biochemical components, such as pharmacokinetic enzymes and molecular targets within a biochemical pathway.
- Traditional statistical analytical methods, such as (logistic) regression, are not suitable for detecting complex gene-gene interactions.

Multifactor dimensionality reduction and classification and regression tree

- Statistical analysis testing for gene-gene interactions can be performed using multifactor dimensionality reduction (MDR) or classification and regression tree (CART) analysis.
- The MDR and CART techniques have been applied successfully to identify genetic classifiers of sunitinib-induced toxicity.

Interpretation and validation of the results

- The MDR and CART techniques both result in genetic classifiers that predict drug response.
- These genetic classifiers must be validated in new patients before they can be used to individualize treatment.

Conclusion and future perspective

- The MDR and CART methods are more rational approaches to individualizing drug treatment when compared with traditional methods.

References

1. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet 2003;361:865-72.
2. Braun MS, Richman SD, Thompson L, *et al.* Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. J Clin Oncol 2009;27:5519-28.
3. Marsh S, Paul J, King CR, *et al.* Pharmacogenetic assessment of toxicity and outcome after platinum plus taxane chemotherapy in ovarian cancer: the Scottish Randomised Trial in Ovarian Cancer. J Clin Oncol 2007;25:4528-35.
4. Grossman I, Sullivan PF, Walley N, *et al.* Genetic determinants of variable metabolism have little impact on the clinical use of leading antipsychotics in the CATIE study. Genet Med 2008;10:720-9.
5. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306-9.
6. Ikediobi ON, Shin J, Nussbaum RL, *et al.* Addressing the challenges of the clinical application of pharmacogenetic testing. Clin Pharmacol Ther 2009;86:28-31.
7. Kooloos WM, Wessels JA, van der Straaten T, Huizinga TW, Guchelaar HJ. Criteria for the selection of single nucleotide polymorphisms in pathway pharmacogenetics: TNF inhibitors as a case study. Drug Discov Today 2009;14:837-44.
8. Wilke RA, Reif DM, Moore JH. Combinatorial pharmacogenetics. Nat Rev Drug Discov 2005;4:911-8.
9. Baker SD, Verweij J, Cusatis GA, *et al.* Pharmacogenetic pathway analysis of docetaxel elimination. Clin Pharmacol Ther 2009;85:155-63.
10. Beretta L, Cappiello F, Moore JH, *et al.* Ability of epistatic interactions of cytokine single-nucleotide polymorphisms to predict susceptibility to disease subsets in systemic sclerosis patients. Arthritis Rheum 2008;59:974-83.
11. Menon R, Velez DR, Simhan H, *et al.* Multilocus interactions at maternal tumor necrosis factor-alpha, tumor necrosis factor receptors, interleukin-6 and interleukin-6 receptor genes predict spontaneous preterm labor in European-American women. Am J Obstet Gynecol 2006;194:1616-24.
12. Tsai CT, Lai LP, Lin JL, *et al.* Renin-angiotensin system gene polymorphisms and atrial fibrillation. Circulation 2004;109:1640-6.
13. Ye Y, Yang H, Grossman HB, *et al.* Genetic variants in cell cycle control pathway confer susceptibility to bladder cancer. Cancer 2008;112:2467-74.
14. Ye Y, Lippman SM, Lee JJ, *et al.* Genetic variations in cell-cycle pathway and the risk of oral premalignant lesions. Cancer 2008;113:2488-95.
15. Gupta M, Chauhan C, Bhatnagar P, *et al.* Genetic susceptibility to schizophrenia: role of dopaminergic pathway gene polymorphisms. Pharmacogenomics 2009;10:277-91.
16. Gordon MA, Gil J, Lu B, *et al.* Genomic profiling associated with recurrence in patients with rectal cancer treated with chemoradiation. Pharmacogenomics 2006;7:67-88.
17. Wilke RA, Moore JH, Burmester JK. Relative impact of CYP3A genotype and concomitant medication on the severity of atorvastatin-induced muscle damage. Pharmacogenet Genomics 2005;15:415-21.
18. Dervieux T, Wessels JA, van der Straaten T, *et al.* Gene-gene interactions in folate and adenosine biosynthesis pathways affect methotrexate efficacy and tolerability in rheumatoid arthritis. Pharmacogenet Genomics 2009;19:935-44.
19. Perez-Perez JM, Candela H, Micol JL. Understanding synergy in genetic interactions. Trends Genet 2009;25:368-76.
20. Moore JH, Gilbert JC, Tsai CT, *et al.* A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 2006;241:252-61.
21. van Erp NP, Eechoute K, van der Veldt AA, *et al.* Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. J Clin Oncol 2009;27:4406-12.
22. Ritchie MD, Motsinger AA. Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies. Pharmacogenomics 2005;6:823-34.
23. Pattin KA, White BC, Barney N, *et al.* A computationally efficient hypothesis testing method for epistasis analysis using multifactor dimensionality reduction. Genet Epidemiol 2009;33:87-94.
24. Moore JH, Ritchie MD. The challenges of whole-genome approaches to common diseases. JAMA 2004;291:1642-3.
25. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.

Websites

101. http://sourceforge.net/projects/mdr/
     Freely available MDR software (accessed February, 2010)
102. https://sourceforge.net/projects/mdr/files/mdrpt/
     Freely available permutation testing software for MDR (accessed February, 2010)