Published in final edited form as:

Pharmacogenet Genomics. 2008 November ; 18(11): 955–965. doi:10.1097/FPC.0b013e32830efdd4.

Germline Genetic variations in drug action pathways predict clinical outcomes in advanced lung cancer treated with platinum-based chemotherapy

Xifeng Wu¹, Charles Lu², Yuanqing Ye¹, Joe Chang³, Hushan Yang¹, Jie Lin¹, Jian Gu¹, Waun Ki Hong², David Stewart², and Margaret R. Spitz¹

¹Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

²Thoracic/Head & Neck Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

³Radiation Oncology (JC), The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030

Abstract

Purpose—Genetic polymorphisms contribute to interindividual variation in drug response. However, a single polymorphism is likely to exhibit a modest effect. Therefore, we applied a pathway-based approach to evaluate the cumulative effect of multiple polymorphisms on clinical outcome of patients with non-small cell lung cancer (NSCLC).

Methods—We genotyped 25 functional polymorphisms in 16 key genes involved in cisplatin metabolism and action and evaluated their associations with overall survival in 229 NSCLC patients receiving first-line cisplatin-based chemotherapy.

Results—Several biologically plausible main effects were identified in individual analysis. More importantly, when 6 polymorphisms in nucleotide excision repair genes were analyzed jointly, a significant trend of reduced risk of death with decreasing number of putative unfavorable genotypes was observed (P for trend <0.001 and log-rank $p < 0.001$). Survival tree analysis revealed potential higher-order gene-gene interactions and categorized subgroups with dramatically different survival experiences, based on distinct genotype profiles. The median survival time was 78.5 months for terminal node 1 in the low-risk group, 15.1 months for terminal node 10 in the medium-risk group, and 6.7 months for terminal node 9 in the high-risk group (log rank $P < 0.001$). We also constructed a prediction hazard model. The area under the curve (AUC) increased from 0.71 (using clinical variables only) to 0.84 (using clinical, epidemiological, and genetic variations from survival tree analysis).

Conclusions—Our results highlight the clinical potential of taking a pathway-based approach and using survival tree analytic approach to identify subgroups of patients with distinctly differing outcomes.

Introduction

The annotation of the human genome provides an opportunity to explore the impact of genetic variation in determining survival differences in non-small cell lung cancer (NSCLC), the leading cause of cancer mortality. Patients with NSCLC are commonly treated with platinum-based chemoradiotherapy and the response rate varies but is generally less than 20% [1].

Significant toxicities that may be lethal are frequently observed. Wider application of cisplatin in NSCLC treatment has been impeded by this intrinsic or acquired resistance [2]. Therefore, the ability to predict therapeutic response in these patients is of immense clinical benefit.

Currently only clinical variables are used to guide treatment decisions with only modest ability to predict overall survival [1]. Molecular signatures derived from global gene expression profiling have shown promise in predicting clinical outcome [3-6], as have pathway-based or genome-wide identification of somatic aberrations using high-density comparative genomic hybridization in tumor tissues [7-9]. However, since these approaches utilize tumor tissues, most of the findings cannot be readily translated into clinical practice due to the difficulty in sample procurement and tumor heterogeneity. Moreover, differences in surgical resection, tissue storage, and experimental procedures, have resulted in non-reproducibility of the findings [10].

The use of germline genetic variants such as single nucleotide polymorphisms (SNPs) is an alternative and complementary approach and has produced promising results [11-13]. The pharmacogenetics of cisplatin in particular, has attracted wide interest. Cisplatin and other platinum agents bind preferentially to DNA. The level of platinum-DNA adducts in the circulation is correlated with clinical outcome and resistance to platinum agents has been linked to enhanced tolerance and repair of DNA damage.

Nucleotide excision repair (NER) is the primary DNA repair pathway responsible for the removal of cisplatin-DNA adducts [14]. Other cisplatin-related pathways include drug uptake, metabolism, and efflux, regulation of cell cycle checkpoints, and apoptosis. Many studies have evaluated the association between common genetic variations in major NER and other genes and cisplatin response, but the results have been inconsistent [15-17]. It is apparent from current literature that individual polymorphism in a single gene would have minimal to modest effect on platinum drug response.

In this study, in an attempt to think beyond the candidate gene approach and identify clinically relevant pharmacogenetic markers, we genotyped 25 potential functional polymorphisms in 16 cisplatin-relevant genes in 229 patients with advanced NSCLC. We then applied several analytic tools to explore the cumulative effects of multiple variants and gene-gene interactions in modulating the survival of cisplatin-treated NSCLC patient.

Methods

Patient characteristics

Subjects in this analysis were newly diagnosed, histologically confirmed, lung cancer patients who had not been previously treated (by radiotherapy and/or chemotherapy) and who were enrolled into an ongoing epidemiologic lung cancer study at The University of Texas M. D. Anderson Cancer Center. From this database of almost 2,000 lung cancer cases, we selected all patients with NSCLC who were staged as IIIB (wet or dry) or IV and who had received first-line cisplatin-based chemotherapy at M. D. Anderson. We further restricted the case series to non Hispanic whites to control for confounding by ethnicity.

Data collection

All subjects signed a consent form and were interviewed using a structured questionnaire to elicit epidemiological data, including demographics, smoking history, alcohol consumption, family history of cancer, medical history, and occupational exposures. At the end of the interview, 40 ml of blood was drawn into coded heparinized tubes. Clinical and follow-up data were abstracted from medical records. The study end point was overall survival. The study was

approved by the institutional review board of The University of Texas M. D. Anderson Cancer Center.

Genotyping

Genomic DNA was extracted from peripheral blood. We selected representative candidate genes involved in pathways relevant to cisplatin action, including drug transport, metabolism, NER, cell cycle control, and apoptosis. The genes involved in cisplatin action are continuously updating, and the CREATE Pharmacogenetic Research Network provides recent knowledge on the major pathways and genes in cisplatin pathway (<http://www.pharmgkb.org/do/serve?objId=PA150642262&objCls=Pathway>). Potential functional SNPs were selected from published association studies and from the dbSNP database of the National Center for Biotechnology Information. We used the Taqman method to genotype SNPs on a 384-well ABI 7900HT Sequence Detection System (Applied Biosystems, Foster city, CA) according to the manufacturer's instructions. Typical amplification mixtures (5 μ L) contained sample DNA (5 ng), 1 $\times$ *Taqman* buffer A, dNTPs (200 μ mol/L), $MgCl_2$ (5 mmol/L), AmpliTaq Gold (0.65 units, Applied Biosystems), each primer (900 nmol/L), and each probe (200 nmol/L). The thermal cycling conditions consisted of one cycle of 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. Previously genotyped samples and water wells were included as positive controls and negative controls, respectively. 5% of all samples were randomly selected to run in duplicate and results were 100% concordant.

Statistical analysis

The chi-square (χ^2) test or Fisher's exact test were applied to compare the distribution of selected demographic and clinical variables by vital status. The Cox proportional hazard model was used to assess the effect of individual SNPs on overall survival, defined as the time from the date of first platinum-based treatment to the date of death or the date of last follow-up. Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated by fitting the Cox model while adjusting for age, sex, clinical stage, performance status, chemotherapy regimen, weight loss and smoking history, where appropriate. Kaplan-Meier curves and log-rank tests were used to assess the differences in survival by individual polymorphisms. We looked at three year survival as well as overall survival and since the results were so similar, only overall survival data are presented. We also performed a pathway-based analysis by counting the number of putative unfavorable genotypes in the NER pathway for those genotypes showing a borderline significance ($p < 0.1$) in the univariate Cox model. Subjects were then trichotomized into low-, medium-, and high-risk groups based on the tertile distribution of the number of unfavorable variants. Using the low-risk group as reference, HRs and 95% CIs were calculated for the medium-risk and high-risk groups using multivariate Cox proportional hazard models adjusting for appropriate variables. STATA software (version 8, STATA Corp., College Station, TX) was used for the above analyses. Survival tree analyses using recursive-partitioning were performed to investigate higher order gene-gene interactions and identify subgroups of individuals at higher risk of death using a modified STREE program (<http://masal.med.yale.edu/stree/>) [18]. STREE is a tree program for survival analysis. The tree starts with the root node which includes all the study subjects and uses a log-rank statistic to select the optimal split that distinguishes patients into better and worse survival. The recursive procedure continues to produce subsequent nodes that are more homogeneous (with respect to survival) than the original node. The final model is a tree-structure with many binary splits, and each terminal node represents a group of patients with different survival pattern based on distinct genotype combinations. We also constructed receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) to evaluate the specificity and sensitivity of predicting death and survival by clinical variables only, and by the combination of clinical, epidemiological, and genetic variables.

Results

Characteristics of patients

This report is based on 229 patients enrolled from 1991 to 2004 with a median follow-up of 12.6 months (Supplement Table 1). At the time of analysis, only 40 (17.5 %) were still alive. There were 135 men (59.0%) and 94 women (41.0%). The majority of the patients (62%) presented with stage IV NSCLC. The histological types were as follows: 129 adenocarcinoma (56.33%), 42 squamous cell carcinoma (18.34%), 36 non-differentiated NSCLC (15.72%), 7 large cell carcinoma (3.06%), and 15 other types (6.55%). Men were slightly overrepresented in the surviving group (63% versus 58%). Surviving patients were nearly 3 years younger at diagnosis (mean age, 55.35 ± 11.86 vs. 58.10 ± 10.88 years), more likely to be never smokers (35% vs. 22%), and reported lighter pack year histories, (26.6 vs 34.6). But all these differences did not achieve statistical significance. Those alive were less likely to be Stage IV (68% vs. 73%), to have a performance status of 2-4 (3 vs. 13%), and to report >10% weight loss (5% vs. 10%). However, only the last comparison was statistically significant, in that greater weight loss was a significant risk predictor for death ($p=0.042$) in this cohort.

Risk associated with individual SNPs

Most of the SNPs in the same genes were not in strong linkage disequilibrium (LD). The pairwise LD coefficients (r^2) were as follows: for XPC, PAT and Lys939Gln, $r^2=0.72$, PAT and Ala499Val, $r^2=0.16$, Lys939Gln and Ala499Val, $r^2=0.22$; for XPD, Asp312Asn and Lys751Gln, $r^2=0.49$; for ERCC6, Met1097Val and Arg1230Pro, $r^2=0.02$; for MDR, Ile1144Ile and Ala892Ser, $r^2=0.49$; for GSTP1, Ile105Val and Ala114Val, $r^2=0.19$; for p53, Pro72Arg and intron 6 (rs1625895), $r^2=0.32$, Pro72Arg and intron 3 (16bp insertion), $r^2=0.18$, intron 6 and intron 3, $r^2=0.68$; for STK15, Phe31Ile and Ile57Val, $r^2=0.04$.

Table 1 summarizes the HRs for overall survival (adjusted for age, sex, stage, performance status, chemotherapy regimen, weight loss and pack-years) for all the investigated SNPs. Among NER genes, patients carrying the rare homozygous *RAD23B* genotype (TT) exhibited significantly poorer survival (HR=2.08, 95% CI, 1.04 to 4.16). *ERCC1* variant-containing genotypes (GT+TT) exhibited significantly improved overall survival (HR=0.68, 95% CI, 0.48-0.95). Kaplan-Meier estimates demonstrated that the median survival time (MST) for individuals with the GT+TT genotypes was 15.8 months compared with MST for the wild-type GG genotype of 12.1 months (Fig. 1A).

We also assumed an additive model and summed the unfavorable genotypes (based on HRs from the main effect analysis that had P values <0.1, Table 1) in the NER pathway. These unfavorable genotypes were *XPA_23* (AG/GG), *XPC_intron9* (-), *XPD_751* (AA/AC), *CCNH* (TT/TC), *RAD23B* (TT), and *ERCC1* (GG). There was a trend of reduced risk of death with decreasing number of unfavorable genotypes. Compared with patients ≥ 5 unfavorable genotypes, the HRs for patients with 3-4, and <3 unfavorable genotypes were 0.69 (0.45-1.04) and 0.27 (0.14-0.52), respectively (P for trend <0.001, Table 1). The MST was 10.9, 12.9 and 26.9 months for patients with ≥ 5 , 3 to 4, and <3 unfavorable genotypes, respectively (log-rank $p<0.001$, Fig 1B).

We also genotyped 13 SNPs from genes in pathways related to cisplatin transport, metabolism, apoptosis and cell cycle control. Only the rare homozygous genotype (GG) of *GSTP1* exon5 polymorphism was associated with a significantly altered survival (HR=1.99, 95% CI 1.14-3.45 for overall survival, Table 1). However, none of the above individual associations remained statistically significant after adjusting for multiple comparisons (data not shown).

Since stage IIIB (dry) patients without pleural effusion have a distinct survival distribution from IIIB (wet) and IV patients, we also analyzed these SNPs in IIIB (wet) and IV patients

only (Supplementary Table 2). The individual associations were generally in the same trend as the combined population, but less significant due to smaller sample size. The HRs for the above mentioned individual SNPs were 1.54 (95% CI, 0.67 to 3.50) for *RAD23B* homozygous variant TT genotype, 0.72 (95% CI, 0.49 -1.05) for the *ERCC1* variant-containing genotypes (GT+TT), and 2.48 (95% CI, 1.35 - 4.52) for the homozygous variant genotype (GG) of *GSTP1*. Again, none of the above individual associations remained statistically significant after adjusting for multiple comparisons (data not shown).

Survival tree analysis for overall survival

NER Genes—Fig. 2A shows the resultant tree structure that incorporated the 12 polymorphisms in the NER pathway. *ERCC1* was identified as the initial split, consistent with the results of the main effect analysis. Fig. 2B summarizes the HRs and 95% CIs for the terminal nodes of the survival tree. Node 2 with individuals exhibiting the longest MST of 23.7 months was defined as the reference node. We then classified the terminal nodes into three categories based on the tertile distribution of the HR estimates: low-risk ($HR < 1$), medium-risk ($1 \leq HR \leq 1.2$), and high-risk group ($HR > 1.2$). Kaplan-Meier estimates are plotted in Fig 2C for three representative nodes selected from each risk group. Terminal node 1 is in the low-risk group with a MST of 22.6 months, terminal node 3 is in the medium-risk group with a MST of 17.7 months, and terminal node 4 is in the high-risk group with a MST of 9.27 months (log rank $P < 0.001$). This illustrates the discriminatory ability of the survival tree to identify subgroups of individuals at distinctive risk levels.

All investigated genes—Fig. 3A illustrates the survival tree combining all the investigated polymorphisms. The top three splits were from NER genes (*ERCC1*, *XPC_PAT*, and *XPB1104*), confirming the importance of the NER gene pathway in the outcome of platinum-based therapy. Fig. 3B lists the HR estimates for each terminal node using terminal node 1 as the reference. The terminal nodes are categorized into three risk groups: low-risk group ($HR < 3$), medium risk group ($HR \sim 4$), and high-risk group ($HR > 4$). Fig. 3C depicts the Kaplan-Meier survival curves for three representative nodes from each risk group. The MST is 78.5 months for terminal node 1 in the low-risk group, 15.1 months for terminal node 10 in the medium-risk group, and 6.7 months for terminal node 9 in the high-risk group (log rank $P < 0.001$). The log rank statistics for the three risk groups identified by survival trees that were built using all investigated genes and NER genes are 52.19 and 15.78, respectively. The corresponding p-values for the log rank statistics are $4.6E-12$ and 0.0004, respectively. This result clearly demonstrated the improved precision of this tree versus the others.

Model discrimination ability

To further illustrate the improvement in the discriminatory ability of the model under different scenarios of variable (i.e., predictor) combinations, we calculated the area under curve statistics (AUC) by sequentially adding established prognostic clinical variables (clinical stage, weight loss, performance status, chemotherapy regimen), epidemiological variables (age, sex, smoking status) and terminal nodes from survival tree for all investigated genes (terminal node 6 and terminal node 12 were combined since terminal node 12 provided a perfect prediction and was dropped by logistic regression in the analysis) (Figure 4). The AUC increased from 0.71 (clinical variables only), to 0.73 (clinical + epidemiologic variables), and to 0.83 (clinical variables, epidemiological variables, genetic variations from survival tree analysis). The difference of AUC between the latter two models (0.83 and 0.73) was 0.10. Results based on 1,000 bootstrapping samples showed that the distribution of the difference of the AUCs has a 95% bias corrected confidence interval 0.05-0.20, which shows statistical significance and suggests the significant improvement of prediction after adding genetic marker to the model.

Discussion

In this study, we evaluated the impact of 25 potentially functional polymorphisms in cisplatin action pathway genes on clinical outcome of 229 advanced stage NSCLC patients receiving first line platinum-based chemotherapy. We found a few suggestive biologically plausible individual associations, but none of which were significant after adjusting for multiple comparison, suggesting the individual effects of polymorphisms on cisplatin outcome, if exist, are minimal to modest. However, our joint analysis of multiple variants showed that there is a significant gene dosage effect, supporting the clinical potential of taking a polygenic analytic approach in predicting clinical outcomes. Furthermore, our exploratory analyses using the survival tree analysis identified subgroups of patients with different risk levels based on combinations of genotypes.

Previous functional studies have shown that inter-individual differences in DNA repair capacity (DRC) affect chemotherapy outcome in NSCLC patients. Using an *in vitro* host cell reactivation assay, elevated DNA repair capacity was associated with intrinsic resistance of lung cancer to chemotherapy [19], and with poor survival in advanced NSCLC patients receiving chemotherapy [20]. In the latter study, the risk of death increased to 1.11 (95% CI = 1.02 to 1.21) for every percentage increase in DRC in NSCLC patients receiving chemotherapy, but elevated DRC was not a risk factor for death in patients who were not treated with chemotherapy. Biologically, chemotherapy drugs (e.g., cisplatin-based regimen in NSCLC patients) mostly execute cytotoxicity through causing DNA damage, and therefore elevated DNA repair capacity would reduce drug efficacy. This disparate role of DNA repair in cancer susceptibility and drug resistance, a double-edged sword, has been well documented [21]. The observation that low expression of ERCC1 was associated with better survival in NSCLC patients receiving platinum-based chemotherapy [22] is consistent with this notion.

Based on the assumption that genetic variations in DNA repair genes affect DNA repair capacity, there have been many studies investigating DNA repair gene polymorphisms and clinical outcomes of cisplatin-based chemotherapy; however, few studies have been reproducible. For example, we observed that individuals with the variant A alleles of the *ERCC1* 8092C>A had a significantly improved survival (HR=0.68, 95%CI, 0.48-0.95). However, Zhou et al. [23] reported that the variant A allele was associated with a 50% increased risk of death in stage III (A+B) and IV patients. They further showed that the increased risk of death reached 2-fold in stage III (A+B) patients, but not significant in stage IV patients in stratified analysis. Another study showed that the variant allele was associated with higher toxicity in NSCLC patients treated with cisplatin [24]. The studies of XPD Asp312Asn and Lys751Gln in cisplatin response and survival of NSCLC patients provide a more revealing example of the inconsistency in pharmacogenetics literature. One study showed that the variant allele of Asp312Asn was associated with a shorter overall survival [25]; one study suggested the homozygous variant genotype of Lys751Gln conferred a reduced risk of death [26]; and five studies showed null results of either SNPs [27-31]. In our study, the homozygous variant genotype of Lys751Gln exhibited a borderline significant protective effect on the risk of death. The irreproducibility of specific polymorphism effects can be attributed to several reasons. The sample sizes were small. Our sample size of 229 is the largest of all the published NSCLC cisplatin pharmacogenetics studies, but is still considered small by advanced NSCLC study standards. Heterogeneity of tumor stage clearly is one of the main reasons. Stage IIIA and IIIB NSCLC patients exhibited different survival patterns and stage IIIB patients are more similar to stage IV patients in survival distribution. Even in stage IIIB patients, those with pleural effusion (IIIB wet) have a distinct survival distribution from those without (IIIB, dry). Among the above mentioned studies, some included stage IIIA, IIIB, and IV patients [25,28], some included IIIB and IV patients [27,30], and others included IIIB (wet) and IV [26,29]. Treatment heterogeneity is another concern. Different combinatorial regimens (e.g., cisplatin plus

gemcitabine or docetaxel) may also modify the association between polymorphisms and clinical outcomes. In the future, large sample size of homogeneous patient population (in terms of tumor stage and treatment regimen) is necessary to produce reliable and consistent data.

The main observation from this study points to the potential of evaluating cumulative effect and gene-gene interaction of multiple variants. The inability to identify individual candidate markers with substantial effect in cisplatin pathways is not surprising, since all of these polymorphisms exhibit at best modest effect on protein function and platinum agents do not target a specific protein. In addition, both DNA repair and GST systems are highly complex and are regulated with redundant or alternative mechanisms. To enhance the predictive power, we exploited a polygenic approach through assessing the cumulative effects of multiple genetic variants on outcome. First advocated by Mohrenweiser et al. [32], such an approach has shown great promise in enhancing the predictive power in cancer risk assessment recently [33,34]. For example, the cumulative effect of 5 confirmed risk alleles in 5 chromosome regions for prostate cancer and family history was compellingly demonstrated that confers an almost 10-fold increased risk of prostate cancer and accounts for 46% of prostate cancer cases in Swedish men [34]. Similar approach has been championed in clinical outcome prediction. One notable success story of this approach is the pharmacogenetics of anticoagulant drug Warfarin. Warfarin targets vitamin K epoxide reductase complex 1 (VKORC1) and is metabolized mainly by CYP2C9. The combination of VKORC1 and CYP2C9 genotypes provides significantly enhanced power to identify warfarin-sensitive patients who would require a lower maintenance dose of the drug than analyzing either gene alone [35]. In cancer pharmacogenetics, several studies, although still exploratory, have provided proof of principle for this approach. Liu et al. reported that higher number of unfavorable genotypes were associated with a worse outcome in patients with advanced melanoma [36]. Stoeckl et al. (37) analyzed the cumulative effect of four genes (XPD, ERCC1, GSTP1, TS) in colorectal cancer patients treated with 5-FU/oxaliplatin, and found that an increasing number of favorable alleles was associated stepwise with a significantly longer survival time. Quintela-Fandino et al. (38) determined the associations between four DNA repair gene SNPs and clinical outcomes in advanced squamous cell carcinoma of head and neck patients with cisplatin-based induction chemotherapy and observed that each variant allele reduced the risk of death by 2.1-fold. We reported that having fewer adverse alleles in 5-FU and NER pathways was associated with longer survival in esophageal cancer [39]. In the present study, we observed a progressively reduced hazard of death with decreasing number of unfavorable genotypes, suggesting a cumulative influence by multiple NER genetic variants on the clinical outcome of cisplatin-treated NSCLC patients. Therefore, future pharmacogenetic studies should emphasize more on this polygenic approach.

In addition to analyze the cumulative effect of multiple variants, we also explored the high-order gene-gene interaction using survival tree analysis. *ERCC1* 8092C>A was selected as the first defining genotype, consistent with the findings in the main effect analysis. The survival of patients in each node with distinct genotype profiles differed significantly (Figs. 2B and 3B), suggesting good discriminative ability of the tree analysis. This approach has been applied in cancer risk prediction [33] and in clinical outcome research [40] and should be applicable in future.

Finally, when we incorporated the genetic information from these survival tree analyses into a multivariate model, we obtained a significant improvement of discriminatory ability over using clinical variables only (The AUC increased from 0.71 to 0.83, Fig. 4). Our clinical prognosis model includes four variables: clinical stage, performance status, chemotherapy regimen, and weight loss. We also added sex, age, and smoking status in the clinical plus host characteristics model. These variables provided the best basic prediction of survival in our patient population. In literature, metastasis (stage IV), performance status, weight loss, sex,

and age have been consistently included in different models to predict survival in advanced NSCLC. For example, Finkelstein et al. [41] included 8 poor survival predictors in their model: lower PS, bone metastasis, male sex, weight loss, subcutaneous metastasis, large-cell carcinoma, shoulder/arm pain, and liver metastasis. Albain et al. [42] found that lower performance status (PS), male sex, age <70 years, were significant independent predictors of poor survival in an analysis of 2,531 locally advanced or metastatic NSCLC patients. Paesmans et al. [43] included 8 variables: stage IV, lower PS, male sex, age, skin metastasis, and a few blood measurements. Hoang et al. [44] included 6 variables that predict poor survival in their model: subcutaneous metastasis, lower performance status, loss of appetite, liver metastasis, $\geq$ four metastatic sites, and no prior lung surgery. It is now widely recognized that treatment response is affected by a complex interplay between genetic and environmental factors, and between host and tumor characteristics. The improvement in the discriminatory ability of the multivariate model compared with that of only clinical variables underscores the importance of considering biomarkers in prediction of clinical outcome. The risk models to predict clinical efficacy and toxicity in NSCLC patients may permit clinicians to identify patients before the start of therapy who are most and least likely to benefit or to develop toxicity from therapy for lung cancer. This could then lead to individualized chemotherapy that maximizes efficacy and quality of life while minimizing toxicity. It may be that some polymorphisms that predict tumor response and toxicity always go hand in hand (i.e. no tumor selectivity). In this case, the polymorphisms will be of clinical use by suggesting the use of other than platinum-based chemotherapy drug. However, for patients with a beneficial tumor response and little toxicity, these models can pre-select patients for platinum-based therapy and the optimal dose. However, considering the pre-selection of the polymorphisms and post-hoc analysis nature in the current study, replications in independent populations are warranted to verify the predictive capability of different polymorphism combinations and the final risk prediction models.

The limitations of this study include the mixed stages of IIIB (wet) and IIIB (dry), a sample size that is not up to current NSCLC standard, and the imperfect selection of genes and SNPs. We selected our genes based on the first version of platinum pathway on PharmGKB database, which is not comprehensive according to current knowledge and missed some newly validated cisplatin response-relevant genes (e.g., CTR1, or SLC31A1). Furthermore, a haplotype-tagging SNP-based approach will provide more complete information of each gene in modulating clinical outcomes. The strength of this study includes a Caucasian patient population that limit the confounding effect of ethnicity. All patients were treated at a single center with excellent follow-up. This relatively homogeneous therapeutic modality limits the confounding of treatment heterogeneity. We performed similar analysis in over 300 patients without receiving first line platinum-based chemotherapy and the results were different from that in patients receiving first line platinum-based chemotherapy (Supplementary Table 3), indicating that these genetic variants are cisplatin drug action predictive markers rather than disease prognosis markers.

In conclusion, in this study, we identified a few suggestive biologically plausible individual associations. We observed that there is a significant gene dosage effect and a gene-gene interaction, supporting the clinical potential of taking a polygenic analytic approach in predicting clinical outcomes. Furthermore, our study suggests that the incorporation of germline genetic variations into risk prediction of clinical outcomes may increase prediction efficiency.

Acknowledgements

Supported by NCI CA 111646, CA 55769, and CA 70907.

References

- Schiller JH, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* 2002;346(2):92–8. [PubMed: 11784875]
- Kartalou M, Essigmann JM. Mechanisms of resistance to cisplatin. *Mutat Res* 2001;478(12):23–43. [PubMed: 11406167]
- Chen HY, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007;356(1):11–20. [PubMed: 17202451]
- Liu R, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007;356(3):217–26. [PubMed: 17229949]
- Potti A, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 2006;355(6):570–80. [PubMed: 16899777]
- Yanaihara N, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006;9(3):189–98. [PubMed: 16530703]
- Kim TM, et al. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res* 2005;11(23):8235–42. [PubMed: 16322280]
- Thomas RK, Weir B, Meyerson M. Genomic approaches to lung cancer. *Clin Cancer Res* 2006;12(14 Pt 2):4384s–4391s. [PubMed: 16857815]
- Janmaat ML, et al. Predictive factors for outcome in a phase II study of gefitinib in second-line treatment of advanced esophageal cancer patients. *J Clin Oncol* 2006;24(10):1612–9. [PubMed: 16575012]
- Alvarado MD, Jensen EH, Yeatman TJ. The potential role of gene expression in the management of primary and metastatic colorectal cancer. *Cancer Control* 2006;13(1):27–31. [PubMed: 16508623]
- Gurubhagavatula S, et al. XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. *J Clin Oncol* 2004;22(13):2594–601. [PubMed: 15173214]
- Bewick MA, Conlon MS, Lafrenie RM. Polymorphisms in XRCC1, XRCC3, and CCND1 and survival after treatment for metastatic breast cancer. *J Clin Oncol* 2006;24(36):5645–51. [PubMed: 17116943]
- Zorbas H, Keppler BK. Cisplatin damage: are DNA repair proteins saviors or traitors to the cell? *Chembiochem* 2005;6(7):1157–66. [PubMed: 15934047]
- Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov* 2005;4(4):307–20. [PubMed: 15789122]
- Rosell R, et al. Influence of genetic markers on survival in non-small cell lung cancer. *Drugs Today (Barc)* 2003;39(10):775–86. [PubMed: 14668933]
- Furuta T, et al. Transcription-coupled nucleotide excision repair as a determinant of cisplatin sensitivity of human cells. *Cancer Res* 2002;62(17):4899–902. [PubMed: 12208738]
- Wei Q, Frazier ML, Levin B. DNA repair: a double-edged sword. *J Natl Cancer Inst* 2000;92(6):440–1. [PubMed: 10716953]
- Zhang, HP.; Singer, B. *Recursive Partitioning in the Health Sciences*. Springer; New York: 1999.
- Zeng-Rong N, Paterson J, Alpert L, Tsao MS, Viallet J, Alaoui-Jamali MA. Elevated DNA repair capacity is associated with intrinsic resistance of lung cancer to chemotherapy. *Cancer Res* 1995;55(21):4760–4. [PubMed: 7585500]
- Bosken CH, Wei Q, Amos CI, Spitz MR. An analysis of DNA repair as a determinant of survival in patients with non-small-cell lung cancer. *J Natl Cancer Inst* 2002;94(14):1091–9. [PubMed: 12122100]
- Wei Q, Frazier ML, Levin B. DNA repair: a double-edged sword. 1. *J Natl Cancer Inst* 2000;92(6):440–1. [PubMed: 10716953]
- Olaussen KA, et al. DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med* 2006;355(10):983–91. [PubMed: 16957145]
- Zhou W, et al. Excision repair cross-complementation group 1 polymorphism predicts overall survival in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Clin Cancer Res* 2004;10(15):4939–43. [PubMed: 15297394]

24. Suk R, et al. Polymorphisms in ERCC1 and grade 3 or 4 toxicity in non-small cell lung cancer patients. *Clin Cancer Res* 2005;11(4):1534–8. [PubMed: 15746057]
25. Gurubhagavatula S, Liu G, Park S, et al. XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. *J Clin Oncol* 2004;22:2594–601. [PubMed: 15173214]
26. De las Peñas R, Sanchez-Ronco M, Alberola V, et al. Polymorphisms in DNA repair genes modulate survival in cisplatin/gemcitabine-treated non-small-cell lung cancer patients. *Ann Oncol* 2006;17:668–75. [PubMed: 16407418]
27. Tibaldi C, Giovannetti E, Vasile E, Mey V, Laan AC, Nannizzi S, Di Marsico R, Antonuzzo A, Orlandini C, Ricciardi S, Del Tacca M, Peters GJ, Falcone A, Danesi R. Correlation of CDA, ERCC1, and XPD polymorphisms with response and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. *Clin Cancer Res* 2008;14(6):1797–803. [PubMed: 18347182]
28. Giachino DF, Ghio P, Regazzoni S, Mandrile G, Novello S, Selvaggi G, Gregori D, DeMarchi M, Scagliotti GV. Prospective assessment of XPD Lys751Gln and XRCC1 Arg399Gln single nucleotide polymorphisms in lung cancer. *Clin Cancer Res* 2007;13(10):2876–81. [PubMed: 17504986]
29. Isla D, Sarries C, Rosell R, et al. Single nucleotide polymorphisms and outcome in docetaxel-cisplatin-treated advanced non-small-cell lung cancer. *Ann Oncol* 2004;15:1194–203. [PubMed: 15277258]
30. Ryu JS, Hong YC, Han HS, et al. Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer* 2004;44:311–6. [PubMed: 15140544]
31. Camps C, Sarries C, Roing B, et al. Assessment of nucleotide excision repair XPD polymorphisms in the peripheral blood of gemcitabine/cisplatin-treated advanced non-small-cell lung cancer patients. *Clin Lung Cancer* 2003;4:237–41. [PubMed: 14624713]
32. Mohrenweiser HW, Wilson DM 3rd, Jones IM. Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. *Mutat Res* 2003;526(12):93–125. [PubMed: 12714187]
33. Wu X, et al. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am J Hum Genet* 2006;78(3):464–79. [PubMed: 16465622]
34. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, Grönberg H. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008 Feb 28;358(9):910–9. [PubMed: 18199855]
35. Sconce EA, Khan TI, Wynne HA, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 2005;106(7):2329–33. [PubMed: 15947090]
36. Liu D, et al. Impact of gene polymorphisms on clinical outcome for stage IV melanoma patients treated with biochemotherapy: an exploratory study. *Clin Cancer Res* 2005;11(3):1237–46. [PubMed: 15709194]
37. Stoecklacher J, Park DJ, Zhang W, et al. A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *Br J Cancer* 2004;91(2):344–54. [PubMed: 15213713]
38. Quintela-Fandino M, Hitt R, Medina PP, et al. DNA-repair gene polymorphisms predict favorable clinical outcome among patients with advanced squamous cell carcinoma of the head and neck treated with cisplatin-based induction chemotherapy. *J Clin Oncol* 2006;24(26):4333–9. [PubMed: 16896002]
39. Wu X, et al. Genetic variations in radiation and chemotherapy drug action pathways predict clinical outcomes in esophageal cancer. *J Clin Oncol* 2006;24(23):3789–98. [PubMed: 16785472]
40. Gordon MA, Gil J, Lu B, et al. Genomic profiling associated with recurrence in patients with rectal cancer treated with chemoradiation. *Pharmacogenomics* 2006;7(1):67–88. [PubMed: 16354126]
41. Finkelstein DM, Ettinger DS, Ruckdeschel JC. Long-term survivors in metastatic non-small-cell lung cancer: an Eastern Cooperative Oncology Group Study. *J Clin Oncol* 1986;4(5):702–9. [PubMed: 3701389]

42. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Survival determinants in extensive-stage non-small-cell lung cancer: the Southwest Oncology Group experience. *J Clin Oncol* 1991;9(9):1618–26. [PubMed: 1651993]
43. Paesmans M, Sculier JP, Libert P, Bureau G, Dabouis G, Thiriaux J, Michel J, Van Cutsem O, Sergysels R, Mommen P, et al. Prognostic factors for survival in advanced non-small-cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation algorithms in 1,052 patients. The European Lung Cancer Working Party. *J Clin Oncol* 1995;13(5):1221–30. [PubMed: 7738625]
44. Hoang T, Xu R, Schiller JH, Bonomi P, Johnson DH. Clinical model to predict survival in chemonaive patients with advanced non-small-cell lung cancer treated with third-generation chemotherapy regimens based on eastern cooperative oncology group data. *J Clin Oncol* 2005;23(1):175–83. [PubMed: 15625371]

Figure 1.

A) Kaplan-Meier curve in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy by *ERCC1* 8092C>A genotype. B) Kaplan-Meier curve in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy by the number of unfavorable genotypes (UFGs) in NER pathway.

Figure 2.
A) Survival tree analysis in NER pathway in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy; B) Cox proportional model in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy based on the survival tree analysis in NER Pathway; C) Kaplan-Meier curves in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy based on survival tree analysis in NER Pathway for selected nodes.

Figure 3.
A) Survival tree analysis in cisplatin pathway in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy; B) Cox Proportional model in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy based on survival tree analysis in cisplatin pathways; C) Kaplan-Meier curves in stages IIIB/IV NSCLC patients treated with first line cisplatin-based chemotherapy based on survival tree analysis in cisplatin pathways for selected nodes.

Figure 4.

ROC curves from various models showing improvement of discrimination ability. True positives (sensitivity) were plotted on the y-axis and false positives (1-specificity) on the x-axis. The area under curve (AUC) measures the discrimination ability of the ROC test to correctly classify those with and without the outcome (disease, response, death, etc). An AUC of 1 represents a perfect discrimination (100% sensitivity and specificity); an AUC of 0.5 represents an accuracy that would be achieved by chance alone (equivalent to tossing a coin); an AUC of 0.83 indicates that there is a 83% likelihood that the model will correctly classify two randomly drawn pairs of patients into dead and survival groups based on the variables included in the model.

Table 1

Cisplatin-relevant Gene Polymorphisms and Clinical Outcomes

NER gene polymorphisms						Other Cisplatin pathway (transporter, metabolism, apoptosis & cell cycle control) gene polymorphisms						
	Alive	Dead	HR ^a	95% CI ^d	P-value		Alive	Dead	HR	95% CI	P-value	
XPA_23 (rs1800975)	40	185				MDR_1144 (rs1045642)	40	182				
	AA	9	40	1			CC	7	50	1		
	AG	13	80	1.34	0.87 to 2.08		0.186	CT	26	80	0.73	0.48 to 1.09
	GG	18	65	1.47	0.94 to 2.30		0.089	TT	7	52	0.86	0.56 to 1.32
	AG/GG			1.4	0.94 to 2.09		0.099	CT/TT			0.78	0.53 to 1.13
P for the trend						P for the trend						
ERCC6_1097 (rs2228526)	40	180				MDR_893 (rs2032582)	40	180				
	AA	25	116	1			GG	13	67	1		
	AG/GG	15	64	1.22	0.86 to 1.73		0.274	GT	21	83	0.92	0.63 to 1.34
							TT	6	30	0.84	0.52 to 1.38	
							GT/TT			0.9	0.63 to 1.28	
P for the trend						P for the trend						
ERCC6_1230 (rs4253211)	40	181				GSTP1 exon5(rs1695)	40	184				
	GG	32	146	1			AA	22	94	1		
	GC	8	35	1.03	0.68 to 1.56		0.886	AG	14	72	1.15	0.80 to 1.64
							GG	4	18	1.99	1.14 to 3.45	
							AG/GG			1.27	0.91 to 1.78	
P for the trend						P for the trend						
XPC_939 (rs2228001)	40	183				GSTP1 exon6 (rs1138272)	40	184				
	AA	14	66	1			CC	32	142	1		
	AC	21	87	0.87	0.61 to 1.25		0.446	CT/TT	8	42	1	0.66 to 1.52
	CC	5	30	0.87	0.52 to 1.45		0.6					
	AC/CC			0.87	0.62 to 1.23		0.426					
P for the trend						P for the trend						

NER gene polymorphisms					Other Cisplatin pathway (transporter, metabolism, apoptosis & cell cycle control) gene polymorphisms					
	Alive	Dead	HR ^a	95% CI ^a	P-value	Alive	Dead	HR	95% CI	P-value
<hr/>										
XPC_PAT (intron 9)	40	180				40	182			
	18	92	1			26	114	1		
	17	63	0.76	0.53 to 1.08	0.125	13	55	1.05	0.73 to 1.50	0.809
	5	25	0.68	0.40 to 1.16	0.158	1	13	1.25	0.66 to 2.39	0.49
			0.74	0.53 to 1.03	0.072			1.08	0.77 to 1.51	0.65
					0.079					0.536
	P for the trend									
<hr/>										
XPC_499 (rs2228000)	35	161				40	184			
	20	92	1			30	135	1		
	12	61	1.06	0.71 to 1.59	0.759	10	49	1.1	0.76 to 1.60	0.603
	3	8	1.15	0.53 to 2.49	0.718					
			1.08	0.73 to 1.58	0.713					
					0.675					0.46
	P for the trend									
<hr/>										
XPD_751 (rs1052559)	40	185				40	185			
	14	80	1			22	108	1		
	20	81	0.95	0.66 to 1.36	0.774	16	56	1.07	0.74 to 1.53	0.734
	6	24	0.65	0.39 to 1.06	0.086	2	21	1.29	0.77 to 2.16	0.337
			0.84	0.60 to 1.18	0.322			1.12	0.81 to 1.55	0.492
					0.115					0.366
	P for the trend									
<hr/>										
XPD_312 (rs1799793)	39	172				40	186			
	17	83	1			27	136	1		
	18	75	0.83	0.58 to 1.19	0.303	13	50	0.99	0.69 to 1.44	0.975
	4	14	0.68	0.37 to 1.27	0.228					
			0.8	0.57 to 1.12	0.196					
					0.158					0.991
	P for the trend									

Other Cisplatin pathway (transporter, metabolism, apoptosis & cell cycle control) gene polymorphisms									
NER gene polymorphisms									
	Alive	Dead	HR ^a	95% CI ^a	P-value	Alive	Dead	HR	P-value
XPGI 104 (rs17655)									
GG	40	183				40	182		
	20	110	1			13	55	1	
GC/CC	20	73	0.75	0.54 to 1.06	0.101	17	83	1.36	0.124
						10	44	1.11	0.654
						GA/AA		1.27	0.202
P for the trend					0.146	P for the trend			
								0.559	
CCNH (rs2266690)									
TT	40	182				40	183		
	26	108	1			5	30	1	
TC	10	62	0.99	0.69 to 1.41	0.955	21	81	1.01	0.955
CC	4	12	0.57	0.29 to 1.11	0.097	14	72	1.2	0.453
TC/CC			0.88	0.63 to 1.23	0.469			1.1	0.679
P for the trend					0.195	P for the trend			
								0.375	
RAD23 (rs1805329)									
CC	40	183				40	182		
	25	115	1			29	117	1	
CT	14	57	1.14	0.80 to 1.64	0.464	10	59	1.24	0.249
TT	1	11	2.08	1.04 to 4.16	0.039	1	6	0.97	0.956
CT/TT			1.25	0.90 to 1.76	0.187			1.21	0.296
P for the trend					0.074	P for the trend			
								0.408	
ERCC1 (rs3212986)									
CC	40	183				40	183		
	17	105	1			23	111	1	
CT/AA	23	78	0.68	0.48 to 0.95	0.026	15	62	0.8	0.196
						2	10	1.43	0.343
						TA/AA		0.84	0.316
P for the trend					0.015	P for the trend			
								0.627	

Other Cisplatin pathway (transporter, metabolism, apoptosis & cell cycle control) gene polymorphisms									
NER gene polymorphisms									
	Alive	Dead	HR ^a	95% CI ^a	P-value		Alive	Dead	
Unfavorable Genotype (Number) ^b						STK15_57 (rs1047972)			
High risk(>=5)	4	38	1			GG	40	181	
Medium risk(3~4)	29	121	0.69	0.45-1.04	0.132	GA/AA	32	137	1
Low risk(0 ~ 2)	7	16	0.27	0.14-0.52	<0.001		8	44	0.86
Log rank test					<0.001				
P for trend					<0.001	P for the trend			0.607

^a Hazard ratio, adjusted for age, sex, clinical stage, performance status, chemotherapy regimen, weight loss, and packyears.

^b Selected unfavorable genotype: *XPA_23*(AG/GG), *XPC_PAT*(-), *XPD_751*(AA/AC), *CCNH* (TT/TC), *RAD23* (TT), and *ERCC1*(CC) (all P<0.1).