

# Adverse Drug Events-based Tumor Stratification for Ovarian Cancer Patients Receiving Platinum Therapy

Chen Wang, PhD<sup>1</sup>, Michael T. Zimmermann, PhD<sup>1</sup>, Christopher G. Chute, MD, Dr. PH<sup>1</sup>,  
Guoqian Jiang, MD, PhD<sup>1</sup>,

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN

## Abstract

*The underlying molecular mechanisms of adverse drug events (ADEs) associated with cancer therapy drugs may overlap with their antineoplastic mechanisms. In a previous study, we developed an ADE-based tumor stratification framework (known as ADEStrata) with a case study of breast cancer patients receiving aromatase inhibitors, and demonstrated that the prediction of per-patient ADE propensity simultaneously identifies high-risk patients experiencing poor outcomes. In this study, we aim to evaluate the ADEStrata framework with a different tumor type and chemotherapy class – ovarian cancer treated with platinum chemotherapeutic drugs. We identified a cohort of ovarian cancer patients receiving cisplatin (a standard platinum therapy) from The Cancer Genome Atlas (TCGA) (n=156). We demonstrated that somatic variant prioritization guided by known ADEs associated with cisplatin could be used to stratify patients treated with cisplatin and uncover tumor subtypes with different clinical outcomes.*

## 1 Introduction

Ovarian cancer is one of leading causes of cancer death among women in the United States. About 70% of patients at diagnosis present with advanced-stage and high-grade serous ovarian cancer (1). Platinum-based chemotherapy is a standard treatment following a cytoreductive surgery, however, approximately 25% of patients develop platinum-resistance within six months and almost all patients with recurrent disease ultimately develop platinum resistance (2). In addition, partly due to the lack of successful treatment strategies, the overall five-year survival rate for high-grade serous ovarian cancer is only 31%. Although several mechanisms have been revealed to contribute to chemotherapy response (3-5), there are no valid clinical or molecular markers that effectively predict the chemotherapy response.

Recently, the cancer research community is actively working on compiling cancer genomic information, and investigating new therapeutic options and tailored treatment for individual patient according to personal tumor genome. A notable example is The Cancer Genome Atlas (TCGA) research network (6, 7). TCGA has released an ovarian cancer dataset containing a large (for genomics) sample size, comprehensive genomic profiles and clinical outcome information (1). The dataset has been utilized to analyze chemotherapeutic response in ovarian cancers in several previous studies (8, 9).

Adverse drug events (ADEs) are a critical factor for selecting cancer therapy options in clinical practice. For example, cisplatin and carboplatin are two commonly used chemotherapy drugs in the treatment of ovarian cancer and are also used to treat other cancer types. In comparison with cisplatin, the greatest benefit of carboplatin is its reduced side effects, particularly the elimination of nephrotoxic effects (4). These side effects have been well documented in the United States Food and Drug Administration (FDA) Structured Product Labels (SPLs). The underlying molecular mechanisms of adverse drug events (ADEs) associated with cancer therapy drugs may also overlap with their antineoplastic mechanisms. Specifically, that the antineoplastic mechanism of action, which kills tumor cells, may be the same mechanism by which healthy cells are damaged leading to toxicity. In a previous study, we developed an ADE-based tumor stratification framework (known as ADEStrata) with a case study of breast cancer patients receiving aromatase inhibitors (10), and demonstrated that the prediction of per-patient ADE propensity simultaneously identifies high-risk patients experiencing poor outcome.

In the present study, we aim to evaluate the feasibility of the ADEStrata framework with a different tumor type and class of therapy – ovarian cancer treated with platinum chemotherapeutic drugs. We first identified a cohort of ovarian cancer patients receiving cisplatin drugs from TCGA, and retrieved somatic mutations for each patient case. We then conducted variant prioritization that was guided by known ADEs of cisplatin represented by Human Phenotype Ontology (HPO) terms. We performed pathway-enrichment analysis and hierarchical clustering, which identified two patient subgroups. We finally conducted a clinical outcome association study to investigate whether the patient subgroups are significantly associated with survival outcome in univariate and multivariate analysis.

## 2 Materials and Methods

## **2.1 Materials**

### **2.1.1 *SIDER: A Side Effect Resource***

The SIDER (SIDE Effect Resource) is a public, computer-readable side effect resource that contains reported adverse drug reactions (11). The information is extracted from public documents and package inserts; in particular, from FDASPLs. In the present study, we utilized the latest version SIDER 2 that was released on October 17, 2012.

### **2.1.2 *HPO: Human Phenotype Ontology***

The HPO project aims to provide a standardized vocabulary of phenotypic abnormalities encountered in human diseases (12). The ontology contains more than 10,000 terms and equivalence mappings to other standard vocabularies such as MedDRA and UMLS. In the present study, we used the latest version of HPO-MedDRA mapping file that is publicly available from the HPO website (13).

### **2.1.3 *eXtasy: A Variant Prioritization Tool***

eXtasy is a variant prioritization pipeline developed at the University of Leuven, for computing the likelihood that a given nonsynonymous single nucleotide variants (nSNVs) is related to a given phenotype (14, 15). The eXtasy pipeline takes a Variant Call File (VCF) and one or more gene prioritization files. Each prioritization file is pre-computed for a specific phenotype (HPO term). In the present study, we downloaded and installed the tool on a local Ubuntu server.

### **2.1.4 *TCGA Data Portal***

TCGA Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA consortium (16). As of September 2014, there are 586 cases of ovarian serous cystadenocarcinoma (OV) with data. In the present study, we utilized the OV clinical data (including clinical drug data and follow-up data) and somatic mutation data through the Open Access data tier.

## **2.2 Methods**

### **2.2.1 *Identifying HPO ADE Terms Relevant to Platinum Drugs***

We first mapped the ADE terms represented in MedDRA UMLS concept unique identifiers (CUIs) from the SIDER 2 database file to the HPO terms using an HPO-MedDRA mapping file produced by HPO development team. Second, we annotated those HPO terms with a flag using the eXtasy HPO term list to indicate whether a HPO-based ADE term can be processed by eXtasy or not. Third, we retrieved those entries (with drug-ADE pairs) using the drug name “cisplatin” and identified a list of ADEs with their HPO term annotations.

### **2.2.2 *Identifying Patient Cohorts by Platinum Drugs and Somatic Mutations from TCGA***

We utilized the clinical drug information file of the OV patients from TCGA data portal through its Open-Access HTTP Directory. The spelling corrections were taken for all variants of the three drugs to maximize the sample size of the patient cases. We then identified a set of patient cases (represented by patient barcodes) that were prescribed for the cisplatin.

We also downloaded the somatic mutation file of the OV patients from TCGA data portal in a Mutation Annotation Format (MAF). The format is a tab-delimited file containing somatic mutations for each patient. As eXtasy requires a VCF file as input, we converted the MAF file into a collection of VCF files. Each VCF file contains somatic mutations for a single patient tumor sample. We combined all VCF files for all cisplatin cases into a single VCF file using the patient barcodes identified in the step above.

### **2.2.3 *Variant Prioritization Using HPO ADE Terms***

As mentioned above, we installed an instance of the eXtasy tool in a local server and ran the tool with a custom Ruby script. The input consists of a VCF file and a set of pre-computed gene prioritization files for those phenotypes represented by the HPO ADE terms of interest. The output is a file with likelihood scores for input variants of impacting an individual HPO term (17). The scores represent the probability that a variant is high-ranking in all different phenotypes comparing against a null distribution of random rankings. To shed some lights on how the variants could potentially affect protein function, we first classified the input variants into three functional impact categories, calling a variant “high” if it is a frameshift, nonsense, nonstop, or splice-site; and “medium” if it is a missense; and “silent” if it is a mutation not causing protein coding changes. And then we analyzed the function of those variants scored by eXtasy for cisplatin-related HPO terms.

### **2.2.4 *Tumor Mutation Stratification and Clinical Outcome Association Studies***

We first selected statistically significant variants based on the eXtasy order statistics (pseudo p-value <0.05). Second, we aggregated genes affected by these prioritized variants across 1,320 canonical pathways collected from the Molecular Signature Database (MSigDB) (18, 19). In order to reduce false discoveries, multiple criteria were applied to further filter out less relevant pathways (binomial distribution p-value >0.05) or pathways containing too few genes (<10 genes). We excluded pathways with less than 10 genes, based on the consideration that small pathways are often subcomponents of larger pathways, and inclusion of them tends to introduce unnecessary

redundancy. Third, we performed hierarchical clustering to highlight pathway-level patterns among cisplatin-treated patients.

We used overall survival (OS) time (years) as a clinical endpoint to measure the outcome of TCGA patients in the identified cohort. We performed both univariate analysis and multivariate cox-regression to assess the association of clusters (produced by hierarchical clustering) with survival. In multivariate analysis, patient age and tumor stage were adjusted for to evaluate the independent outcome-prediction contribution of found tumor cluster. We also analyzed the distribution of patient age and tumor stage in the clusters identified.

### 3 Results

In total, we identified a list of cisplatin-induced ADEs represented in 95 unique HPO Ids. Of them, 73 HPO Ids (76.8%) are covered in eXtasy tool. Table 1 shows a list of such ADEs relevant to renal toxicity.

**Table 1.** A list of cisplatin-induced ADEs relevant to renal toxicity represented in HPO terms.

MedDRA UMLS CUI	MedDRA Label	HPO Id	HPO Label	eXtasy
C0341697	Renal impairment	HP:0000082	Abnormality of renal physiology	YES
C0740394	Hyperuricaemia	HP:0002149	Hyperuricemia	YES
C0235416	Blood uric acid increased	HP:0002149	Hyperuricemia	YES
C1565489	Insufficiency renal	HP:0000083	Renal failure	YES
C0035078	Renal failure	HP:0000083	Renal failure	YES
C0020625	Hyponatraemia	HP:0002902	Hyponatremia	YES
C0595916	Nephropathy toxic	HP:0000112	Nephropathy	YES
C0020598	Hypocalcaemia	HP:0002901	Hypocalcemia	YES
C0151723	Hypomagnesaemia	HP:0002917	Hypomagnesemia	YES
C0020621	Hypokalaemia	HP:0002900	Hypokaliemia	YES
C0151747	Renal tubular disorder	HP:0000091	Abnormality of the renal tubule	YES
C1287298	Urine output	HP:0011036	Abnormality of renal excretion	YES
C0032617	Polyuria	HP:0000103	Polyuria	YES

We were able to identify a cohort of 156 OV patients receiving cisplatin treatment from TCGA OV clinical drug data. Of them, 92 OV patients had somatic mutations identified from OV somatic mutation data.

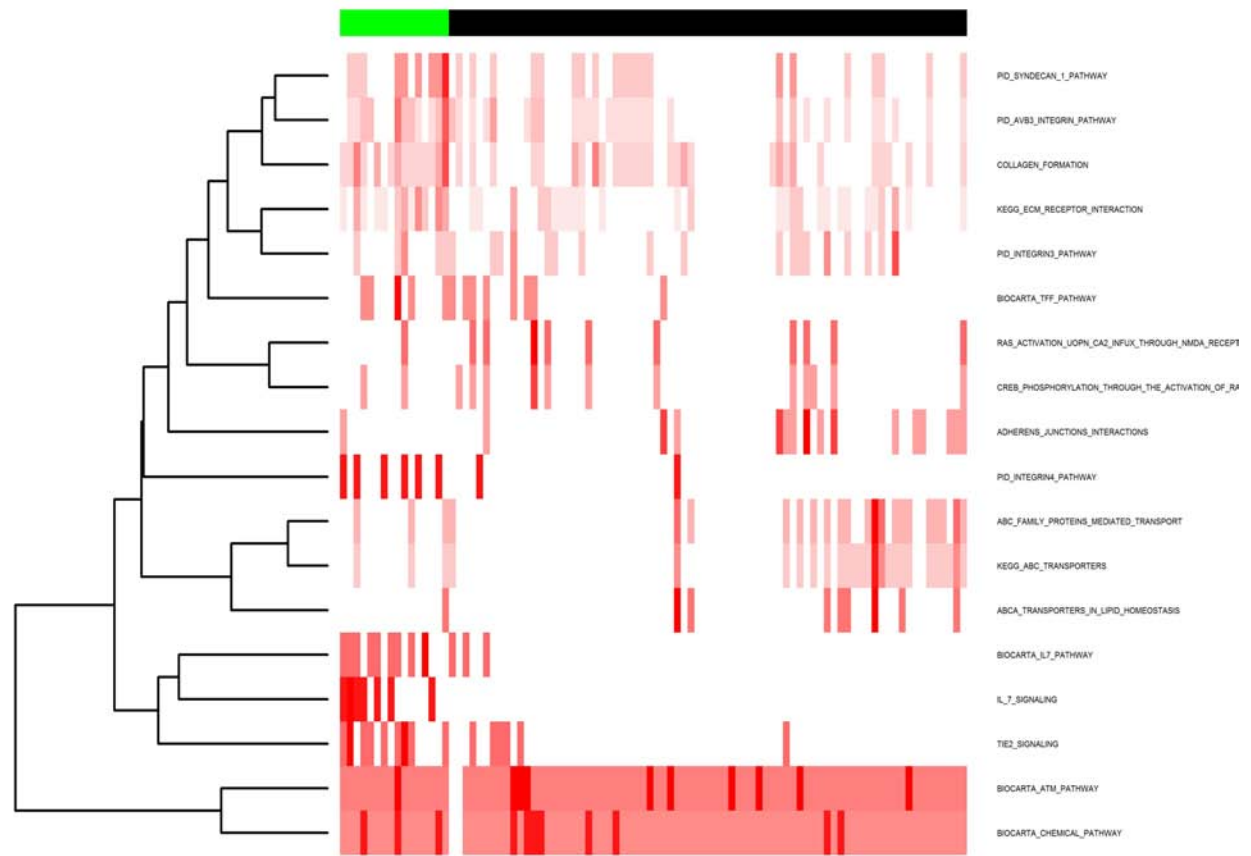
The eXtasy program ignores silent variants. Of the remaining variants, 12% are of high impact (see section 2.2.3) and almost assuredly affect the normal physiologic function of the affected gene. Of the variants scored by eXtasy for cisplatin-related HPO terms, 40% are highly conserved among placental mammals. Because of lack of conservation at many variant sites, approximately 60% cannot be evaluated with common prioritization tools such as SIFT or PolyPhen2. Of those that are evaluable, both SIFT and PolyPhen2 predict 60% of them as deleterious (predictions are 76% concordant). Variants were prioritized for each patient across the ADE phenotypes represented by 73 HPO terms, producing aggregate prioritization scores (max and order statistics).

By hierarchical clustering, 2 distinct patient clusters, organized by pathways (affected by prioritized variants), were identified and are displayed in Figure 1 containing 16 and 76 patients each. Table 2 shows the results of the univariate and multivariate cox-regression analysis for the three clusters. We found that Cluster 2 has a relatively large number of patients (n=76), and is significantly association with poorer survival time in both univariate and multivariate analysis. Table 3 shows the distribution of age and stage in the 2 clusters identified. There is no significant association between the 3 clusters and age/stage, although we noticed that Cluster 2 is enriched with more Stage IIIC and Grade 3 patient cases. Figure 2 shows a Kaplan-Meier plot of survival time for the 2 clusters, derived from our pathway-level analysis, indicating Cluster 2 had the worse survival outcome associated.

### 4 Discussion

While TCGA catalogs a large number of OV samples, sample size for individual chemotherapies may be small. Thus, we focus first on the most common chemotherapy regimen so that the subgroup of interest is still reasonably large. In our previous study we considered patients receiving aromatase inhibitors (10). Aromatase inhibitors block conversion of precursor hormones to estradiol, effectively turning off the growth signal for estrogen-dependent tumors. Evidence exists for tumor addiction; that loss of this dependent growth signal leads to apoptosis. The healthy tissues most likely to be affected by this treatment are those who routinely use the aromatase enzyme or estrogen signaling in their normal physiology. In this study, we consider a platinum-based therapy whose mechanism of action is to nonspecifically damage DNA. Any cell could be affected. The tissues most affected are those who are quickly growing and have a greater fraction of their DNA accessible. These include the cancer itself,

but also hematologic stem cells and those of the digestive tract. The mechanistic link to the studied ADEs is clearer – kidneys become compromised due to higher blood protein levels and blood cells cannot be replaced as quickly. The therapy’s molecular mechanism is responsible for the ADEs considered. The rationale behind nonspecific chemotherapies, such as cisplatin, is to damage tumor cells more than healthy cells, but damage to both is expected.



**Figure 1.** An ordered heatmap showing pathway-level clustering of 92 patients treated with cisplatin across ADE relevant variants. The color of heatmap from white to red indicates low to high percentages (0% to 100%) of genes affected by ADE relevant variants. Column color-bar on top of the heatmap indicates two clusters of samples: Cluster 1 (green) and Cluster 2 (black). Note that the number of the patients (n=92) with pathway enrichment is less than total number of the identified cohort (n=156) is because not all patients have prioritized variants listed.

**Table 2.** The univariate and multivariate cox-regression analysis results of cluster labels. In multivariate analysis, patient diagnosis age, tumor-grade and tumor-stage were adjusted for to determine the independent contribution of cluster membership. HR denotes hazard ratio; \* denotes p<0.05.

Univariate analysis	p-value	HR [95% CI]
Cluster-2	0.019*	3.16 [1.21, 8.30]
Multivariate analysis		
Cluster-2	0.013*	3.47 [1.30, 9.23]
Diagnosis age	0.67	0.99 [0.96, 1.02]
Grade-2	0.18	0.30 [0.05, 1.71]
Grade-3	0.41	0.53 [0.13, 2.34]
Stage-IIIC	0.93	0.96 [0.39, 2.36]
Stage-IV	0.78	0.85 [0.28, 2.61]

**Table 3.** The distribution of age tumor-grade, and tumor-stage in the two clusters identified. # p-value for age vs. cluster association was computed using ANOVA test; p-value for stage/grade vs. cluster association was computed using Fisher’s exact test.

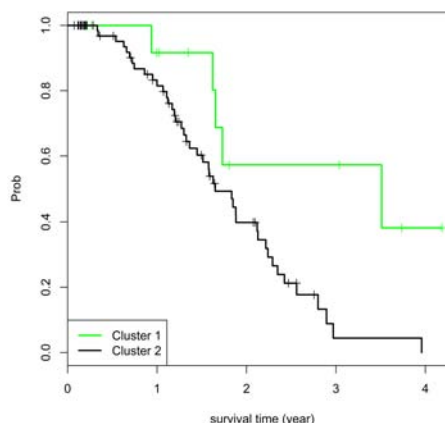
	Cluster-1 (n=16)	Cluster-2 (n=76)	p-value#
Diagnosis age	54.3 [47.1, 53.0, 64.4]	56.2 [49.1, 56.2, 61.3]	0.49
Mean [Q1, median, Q3]			
Stage (case number)			1
II, IIIA or IIIB	2	8	
IIIC	12	56	
IV	2	12	
Grade (case number)			1
Missing	0	3	
G2	1	8	
G3	15	65	

A logical extension of our current methodology would be to independently predict ADEs given germline or somatic variants. High propensity of ADEs from germline alone would predict high toxicity, while high ADE propensity from somatic variants would point to high efficacy. In a given patient, the ideal

situation would be a prediction of low toxicity and high efficacy, while prediction of high toxicity and low efficacy may be a contra-indication for the therapy. An important implication of our findings in this study is that cisplatin could be more toxic than carboplatin but for a subset of patients it could be more effective. We will pursue retrospective validation of this methodology with the long term goal of aiding clinical decision making in personalized cancer treatment.

## 5 Conclusion

In summary, we evaluated the feasibility of ADEStrata framework with a different tumor type and chemotherapy class – ovarian cancer treated with platinum chemotherapeutic drugs. We demonstrated that somatic variant prioritization guided by known ADEs associated with cisplatin could be used to stratify patients treated with cisplatin and uncover tumor subtypes with different clinical outcomes. In the future, we plan to evaluate and validate our approach by incorporating more data types (e.g., germline variants), and investigate the generalization of the method in other tumor types.



**Figure 2.** Kaplan-Meier plot of survival time for patients in 2 pathway-level clusters.

## Reference

- 1 Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011 Jun 30;**474**(7353):609-15.
- 2 Pujade-Lauraine E, Hilpert F, Weber B, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014 May 1;**32**(13):1302-8.
- 3 Ow GS, Ivshina AV, Fuentes G, Kuznetsov VA. Identification of two poorly prognosed ovarian carcinoma subtypes associated with CHEK2 germ-line mutation and non-CHEK2 somatic mutation gene signatures. *Cell Cycle*. 2014 Jul 15;**13**(14):2262-80.
- 4 Dasari S, Bernard Tchounwou P. Cisplatin in cancer therapy: Molecular mechanisms of action. *European journal of pharmacology*. 2014 Oct 5;**740C**:364-78.
- 5 Bowden NA. Nucleotide excision repair: why is it not used to predict response to platinum-based chemotherapy? *Cancer letters*. 2014 May 1;**346**(2):163-71.
- 6 The Cancer Genome Atals. [cited February 17, 2014]; Available from: <http://cancergenome.nih.gov/>
- 7 Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*. 2013 Oct;**45**(10):1113-20.
- 8 Bosquet JG, Marchion DC, Chon H, Lancaster JM, Chanock S. Analysis of chemotherapeutic response in ovarian cancers using publicly available high-throughput data. *Cancer research*. 2014 Jul 15;**74**(14):3902-12.
- 9 Hsu FH, Serpedin E, Hsiao TH, Bishop AJ, Dougherty ER, Chen Y. Reducing confounding and suppression effects in TCGA data: an integrated analysis of chemotherapy response in ovarian cancer. *BMC genomics*. 2012;**13 Suppl 6**:S13.
- 10 Wang C, Zimmermann MT, Prodduturi N, Chute CG, Jiang G. Adverse Drug Event-based Stratification of Tumor Mutations: A Case Study of Breast Cancer Patients Receiving Aromatase Inhibitors. *AMIA Annual Symposium*. Wahsington, DC: (in press); 2014.
- 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.
- 12 Kohler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic acids research*. 2014 Jan 1;**42**(1):D966-74.
- 13 The Human Phenotype Ontology URL. [cited February 17, 2014]; Available from: <http://www.human-phenotype-ontology.org/>
- 14 Sifrim A, Popovic D, Tranchevent LC, et al. eXtasy: variant prioritization by genomic data fusion. *Nature methods*. 2013 Nov;**10**(11):1083-4.
- 15 eXtasy URL. [cited February 15, 2014]; Available from: <http://homes.esat.kuleuven.be/~bioiuser/eXtasy/>
- 16 TCGA Data Portal. [cited February 17, 2014]; Available from: <https://tcga-data.nci.nih.gov/tcga/>
- 17 Aerts S, Lambrechts D, Maity S, et al. Gene prioritization through genomic data fusion. *Nature biotechnology*. 2006 May;**24**(5):537-44.
- 18 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011 Jun 15;**27**(12):1739-40.
- 19 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 Oct 25;**102**(43):15545-50.