Title	:	Genetics Reporting and Analysis Plan for PGx7610: Genetic Evaluation of Hepatotoxicity in Pazopanib Studies
Compound Number	:	GW786034
Effective Date	:	31-JUL-2014

Description :

This Genetics Reporting and Analysis Plan describes planned analyses for eTrack study ID 201761 (PGx7610), which is a confirmatory study of association between genetic variants and pazopanib-related hepatotoxicity. Genetic variants to be tested were previously identified in study VEG117365 (PGx6652) as being associated with alanine aminotransferase (ALT) elevation in pazopanib-treated patients. This study will use an independent sample and data from pazopanib-treated patients enrolled in 23 clinical studies. The primary objective of this analysis is to evaluate genetic associations between *HLA-B**57:01 and ALT elevation in pazopanib-treated subjects. The secondary objective is to evaluate genetic association between 16 other prespecified common genetic variants and ALT elevation in pazopanib treated subjects.

Subject :

Pazopanib (GW786034), PGx7610, pharmacogenetics, PGx, hepatotoxicity, drug-induced liver injury, DILI, safety, human leukocyte antigen, HLA, cancer, alanine aminotransferase, ALT, transaminase

Author's Name and Functional Area:

[redacted] Senior Scientific Investigator (Statistical Genetics)	31-JUL-2014
[redacted] Investigator (Genetics)	31-JUL-2014

Approved by:

[redacted] Director (Statistical Genetics)	31-JUL-2014
[redacted] Oncology Therapy Area Head (Genetics)	31-JUL-2014

TABLE OF CONTENTS

PAGE

1. GENETIC REPORTING & ANALYSIS PLAN SUMMARY 3

2. SUMMARY OF KEY INFORMATION 4
   2.1. Introduction and rationale 4
   2.2. Study Objective(s) and Endpoint(s) 6
   2.3. Study Design 7
   2.4. Statistical Hypotheses 7

3. SAMPLE SIZE CONSIDERATIONS AND POWER ESTIMATES FOR THE PGX ANALYSIS 7

4. GENETIC ANALYSIS POPULATIONS 9

5. CONSIDERATIONS FOR DATA ANALYSES 10

6. DATA HANDLING CONVENTIONS 11

7. PHARMACOGENETIC ANALYSES 11
   7.1. Primary Analyses 11
   7.2. Secondary Analyses 12
   7.3. Other Exploratory Analyses 13
   7.4. General Pharmacogenetic Analysis Conventions 13

8. REFERENCES 14

9. APPENDICIES 15
   9.1. APPENDIX 1: Data Display Standards & Handling Conventions 16
   9.2. APPENDIX 2: Derived and Transformed Data 16
   9.3. APPENDIX 3: Premature Withdrawals & Handling of Missing Data 16
      9.3.1. Premature Withdrawals 16
      9.3.2. Handling of Missing Genetic Data 17
   9.4. APPENDIX 4: Genotype/Subject Quality Control 17
      9.4.1. Subject Quality Control 17
      9.4.2. Genotype Quality Control 17
   9.5. APPENDIX 5: Multiple Comparisons & Multiplicity 18
      9.5.1. Candidate variant analysis 18
   9.6. APPENDIX 6: Hardy-Weinberg (HW) Analysis 18
   9.7. APPENDIX 7: Linkage Disequilibrium Analysis 18
   9.8. APPENDIX 8: Characterizing Ancestry Using Principal Components Analysis 18
   9.9. APPENDIX 9: Estimation of Heritability 18
   9.10. APPENDIX 10: Genotype Imputation 19
   9.11. APPENDIX 11: Abbreviations & Trade Marks 19
      9.11.1. Abbreviations 19
      9.11.2. Trademarks 19

1. GENETIC REPORTING & ANALYSIS PLAN SUMMARY

RAP Area	Description
Purpose	• Planned analyses for 201761 (PGx7610), which is a confirmatory study of association between genetic variants and pazopanib-related hepatotoxicity
Primary Objective / Endpoint	• To evaluate genetic associations between *HLA-B**57:01 and ALT elevation in pazopanib-treated subjects from 23 clinical studies
Study Design	• Retrospective non-interventional pharmacogenetic study
Planned Analyses	• Evaluate genetic associations between *HLA-B**57:01 and 16 pre-specified SNPs and ALT elevation in pazopanib-treated subjects from 23 clinical trials
Analysis Populations	• All pazopanib-treated, as monotherapy or in combination, subjects from 23 clinical trials • Pazopanib monotherapy subjects from 23 clinical trials
Hypothesis	• *HLA-B**57:01 and 16 SNPs, previously identified as potentially associated with ALT elevation in pazopanib-treated subjects, may confirm as being associated with ALT elevation in an independent sample.
Primary Analyses	• Evaluate carriage of a single haplotype (*HLA-B**57:01) in all pazopanib-treated subjects from 23 clinical trials for maximum on-treatment ALT using a one-tailed test
Secondary Analyses	• Test association between carriage of the *HLA-B**57:01 allele, and secondary measures of ALT elevation, in patients treated with pazopanib • Test association between genotypes at 16 pre-specified SNPs, and ALT elevation, in patients treated with pazopanib

2. SUMMARY OF KEY INFORMATION

2.1. Introduction and rationale

The rationale for seeking genetic predictors of alanine aminotransferase (ALT) elevation in pazopanib-treated patients has been described previously (Xu, 2011, VEG117365 [PGx6652] RAP).

In a previous exploratory study that used data from 8 clinical studies (VEG117365, PGx6652), associations were observed between ALT elevation in pazopanib-treated patients, and carriage of the *HLA-B**57:01 allele ($P<5x10^{-4}$; significant after adjusting for 92 common *HLA* alleles tested). In the same exploratory study, genome-wide analyses identified suggestive associations between ALT elevation in pazopanib-treated patients, and SNP genotypes (16 index SNPs with $P<5x10^{-7}$ and minor allele frequency [MAF]≥1%).

This is a confirmatory study that tests the same associated variants in an independent sample of pazopanib-treated patients. Patients enrolled in any of the following 23 clinical studies, who were exposed to at least one dose of pazopanib, and who gave consent and a sample for genetic analyses, will be analysed. The 23 clinical studies are listed in Table 1. It is expected that data from a total of approximately N=1080 patients will be analysed although the exact number will not be known until the data are analysed.

Table 1 Clinical studies included in analysis

Study ID	Phase / Indication	Study Description	Estimated Pazopanib treated with PGx sample
HYT109091	I / Solid Tumours	A Phase I, Open-label, Study of the Safety, Tolerability, and Pharmacokinetics of Two Schedules of Oral Topotecan in Combination with Pazopanib in Subjects with Advanced Solid Tumors	64
VEG10006	I / Solid Tumors	An Open-Label Safety, Pharmacokinetic and Pharmacodynamic Study of Multiple Doses of GW786034 and Lapatinib Concomitantly Administered in Cancer Patients	50
VEG10007	I / Solid Tumors	A Multi-centre, Open-Label, Multiple-probe Drug Interaction Study to Determine the Effects of GW786034 on Metabolism of Cytochrome P450 probe Drugs in Patients with Solid tumors	23
VEG102857	I/II / GBM	Phase I and II, Open-Label, Multi-Center Trials of Pazopanib in Combination with Lapatinib in Adult Patients with Relapsed Malignant Glioma	42
VEG104450	II / Ovarian	A Phase II, Open-Label Study Evaluating the Effect of GW786034 in Subjects with Ovarian Cancer	27

Study ID	Phase / Indication	Study Description	Estimated Pazopanib treated with PGx sample
VEG105281	II / Cervical	A Phase II, Open-Label, Randomized, Multicenter Trial of Pazopanib (GW786034) in Combination with Lapatinib (GW572016) Compared to Pazopanib Monotherapy and Lapatinib Monotherapy in Subjects with FIGO Stage IVB or Recurrent or Persistent Cervical Cancer with Zero or One Prior Chemotherapy Regimen for Advanced/Recurrent Disease	140
VEG105290	II / Lung	A Phase II Open-Label Multicenter Study to Evaluate the Safety and Efficacy of Pazopanib (GW786034) as Neoadjuvant Therapy in Treatment-Naïve Subjects with Stage IA, IB, IIA or IIB (to T2) Resectable Non-Small Cell Lung Cancer (NSCLC)	35
VEG105424	I / CRC	An open-label pharmacokinetic study of the safety and tolerability of pazopanib in combination with FOLFOX 6 or CapeOx in subjects with colorectal cancer	31
VEG105427	I / Breast Cancer	A Phase I, Open-Label, Study of the Safety, Tolerability, and Pharmacokinetics of Pazopanib in Combination with Paclitaxel on a Weekly Schedule for Three Consecutive Weeks of a 28-Day Cycle, Paclitaxel and Carboplatin on an Every 21 Days Schedule and Lapatinib and Paclitaxel on a Weekly Schedule for three Consecutive Weeks of a 28- Day Cycle	60
VEG107200	I / Hepatocellular Cancer	A Phase I, Open-Label, Dose Escalation, Multi-Center Study of pazopanib (GW786034) in Adults Subjects with Hepatocellular Cancer	14
VEG109599	I / Solid Tumors	A Phase I, Open-label, Study of the Safety, Tolerability and Pharmacokinetics of Pazopanib in Combination with Gemcitabine and Gemcitabine plus Cisplatin for Advanced Solid Tumors	22
VEG109603	I / Solid Tumors	An Open-Label, Safety, Pharmacokinetic, and Pharmacodynamic Dose Escalation Phase Ib Study of Pazopanib in Combination with Epirubicin or Doxorubicin in Subjects with Advance Solid Tumors	30
VEG109607	I / Solid Tumors	A Phase I Study of Pazopanib in Combination with Either Erlotinib or Pemetrexed in Patients with Advanced Solid Tumors	52
VEG109609	II / NSCLC	A Phase II, Non-randomized, Multi-center Study to Evaluate the Efficacy and Safety of Pazopanib (GW786034) in Subjects with Advanced Non-Small Cell Lung Cancer	14
VEG109693	I / Solid Tumors	A Phase I, Open-Label, Multiple Dose of Pazopanib Alone and In Combination with Lapatinib in Japanese Patients with Solid Tumors	29
VEG110190	II / Gynecological	A Phase I/II, Open-Label, Multicenter, Two-Arm, Feasibility Study of Pazopanib, Carboplatin, and Paclitaxel in Women with Newly Diagnosed, Previously Untreated, Gynecological Tumors	12
VEG110264	II / Breast Cancer	A Phase II Clinical Trial of Four Cycles of Doxorubicin and Cyclophosphamide Followed by Weekly Paclitaxel Given Concurrently with Pazopanib as Neoadjuvant Therapy for Women with Locally Advanced Breast Cancer Followed by Postoperative Pazopanib	40
VEG111109	II / NSCLC	An open-label, multicenter, phase I/II study of pazopanib in combination with paclitaxel in first-line treatment of subjects with stage IIIBwet/IV non-small cell lung cancer	27
VEG111128	II / NSCLC	An open-label, multicentre, randomised phase II study of pazopanib in combination with pemetrexed in first-line treatment of subjects with predominantly non-squamous cell stage IIIBwet/IV non-small cell lung cancer	69

Study ID	Phase / Indication	Study Description	Estimated Pazopanib treated with PGx sample
VEG113046	III / RCC	PISCES: Pazopanib versus sunitinib patient preference study in treatment naïve metastatic renal cell carcinoma	160
VEG108838	II / IBC	A Randomized, Multicenter, Phase III Study Comparing the Combination of Pazopanib and Lapatinib versus Lapatinib Monotherapy in Patients with ErbB2 over-expressing Inflammatory Breast Cancer	70
VEG108925	I / CRC	Phase I study of Safety and Pharmacokinetics of pazopanib in combination with Cetuximab and irinotecan in patients with colorectal cancer	19
VEG20007	II / Breast Cancer	A Phase II Open-Lable, Randomized, Multicenter Trial of GW786034 (Pazopanib) in Combination with Lapatinib (GW572016) compared to Lapatinib Alone as First Line Therapy in Subjects with Advanced or Metastatic Breast Cancer with ErbB2 Fluorescence In Situ Hybridization (FISH) Positive Tumors	50

2.2. Study Objective(s) and Endpoint(s)

Objectives	Endpoints
Primary	
• Test association between carriage of the *HLA-B**57:01 allele and ALT elevation in patients treated with pazopanib	• Maximum on-treatment ALT (ULN)
Secondary	
• Test association between carriage of the *HLA-B**57:01 allele, and secondary measures of ALT elevation, in patients treated with pazopanib	• Ever vs. never on-treatment ALT>5xULN • Ever vs. never on-treatment ALT>3xULN • Time until first on-treatment ALT>5xULN • Time until first on-treatment ALT>3xULN
• Test association between genotypes at 16 pre-specified SNPs, and ALT elevation, in patients treated with pazopanib	• Maximum on-treatment ALT (ULN) • Ever vs never on-treatment ALT>5xULN • Ever vs never on-treatment ALT>3xULN • Time until first on-treatment ALT>5xULN • Time until first on-treatment ALT>3xULN

Endpoint definitions
• On-treatment ALT measures will be defined as all measures taken between the day after pazopanib treatment initiation, and 28 days after the last dose of pazopanib received (inclusive). • Each ALT measure will be divided by the laboratory- or institution-specific upper limit of normal (ULN) to obtain a measure in ULN units. If no ULN is available the measure will be treated as missing. • **Maximum on-treatment ALT (ULN),** will be the maximum over all on-treatment ALT measures for each patient (in ULN units). Patients with no non-missing on-treatment ALT measures will be excluded from all analyses for this endpoint. • **Ever vs never endpoints** will be defined for two thresholds, on-treatment ALT>5xULN (CTCAE grade 3+) and on-treatment ALT>3xULN (CTCAE grade 2+). All patients in the analysis population will be included for this endpoint. • **Time until first event endpoints** will be defined for two thresholds, on-treatment ALT>5xULN and on-treatment ALT>3xULN. Event times will be measured from initiation of pazopanib treatment until the first threshold-exceeding measure. Subjects with no on-therapy measure of ALT>ULN (no CTCAE grade 1+) will be censored at the end of the on-therapy window. Subjects with maximum on-therapy ALT measure >ULN but not exceeding the threshold (>5xULN or >3xULN) will be excluded from time to event analyses.

2.3. Study Design

This is an observational pharmacogenetic (PGx) study using data from patients enrolled in 23 GSK clinical studies, who were treated with at least one dose of pazopanib, and who provided optional consent and a sample for pharmacogenetic analyses.
The clinical studies to be included are listed in Table 1.

2.4. Statistical Hypotheses

The *HLA-B**57:01 allele, and 16 index SNPs, all previously identified as potentially associated with ALT elevation in pazopanib-treated subjects, will be tested for association with ALT elevation in an independent sample of pazopanib-treated subjects.

3. SAMPLE SIZE CONSIDERATIONS AND POWER ESTIMATES FOR THE PGX ANALYSIS

It is expected that genetic and clinical data will be available for a total of approximately N≈1080 pazopanib-treated patients (Table 1; the PGx pazopanib population, Section 4). Of these, less than one-third (N≈323) were enrolled in clinical studies (or in study arms) where there was no protocol-specified combination therapy for the treatment of cancer (the PGx pazopanib monotherapy population, Section 4), and the remainder (N≈757) were enrolled in studies (or in study arms) with protocol-specified combination therapy for cancer (the PGx pazopanib combination therapy population, Section 4). The previous exploratory analysis (VEG117365, PGx6652), in which the genetic associations

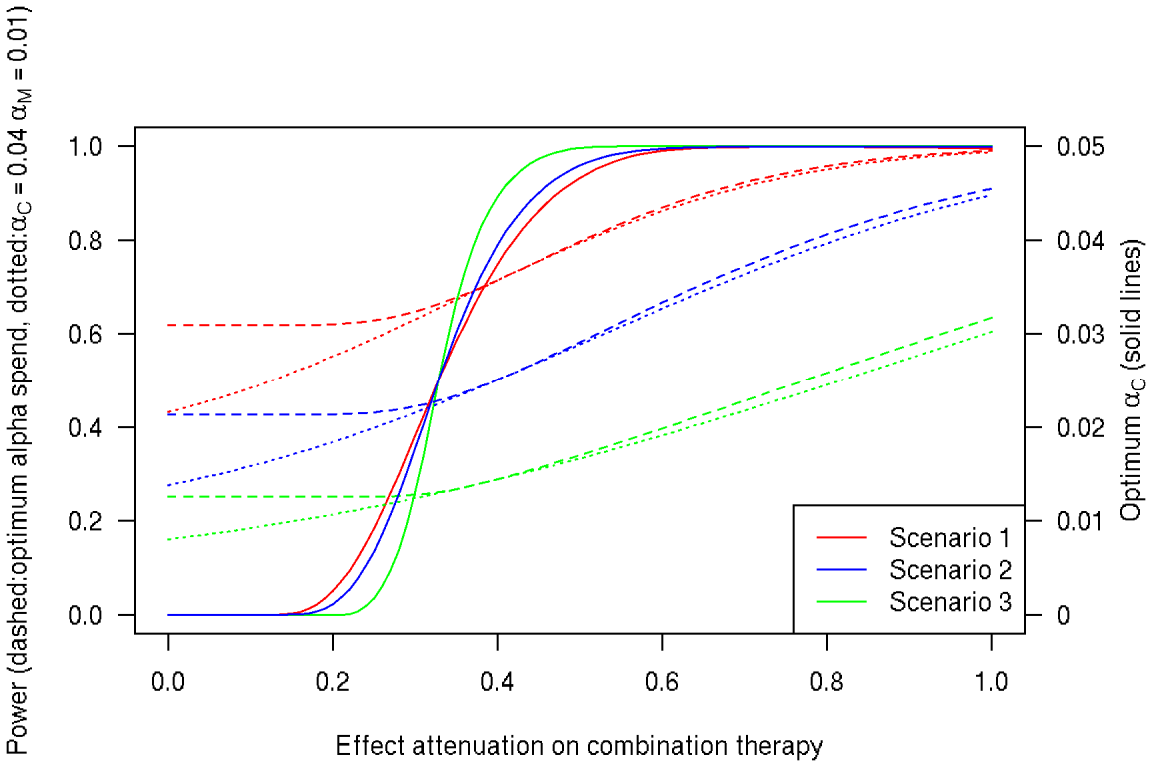
to be tested in the present analysis were discovered, analysed only patients on pazopanib monotherapy (N=1228). Because there are no exploratory PGx data for hepatotoxicity from subjects on pazopanib combination therapy, the optimal strategy for the present study was unclear.

It is biologically plausible that genetic effect sizes in the PGx pazopanib combination therapy population could either be less than, or greater than, genetic effect sizes in the PGx pazopanib monotherapy population. Analysing only the PGx pazopanib monotherapy population could provide stricter sense replication of the associations from the previous exploratory study, but would likely be underpowered (N≈323 for replication compared with N=1228 in discovery). Analysing the PGx pazopanib population (monotherapy and combination therapy combined, N≈1080) would be better powered if there are roughly similar genetic effect sizes in patients on monotherapy and on combination therapy. Furthermore, successful replication in a more heterogenous patient sample would provide greater evidence of robustness of a genetic association, which is important when considering utility for decision making in real world healthcare settings.

Integrating over posterior distribution for effect size estimated in the exploratory discovery study for *HLA-B**57:01 (95% credible interval 1.23—1.89 fold on maximum on-treatment ALT), assuming similar allele frequency and endpoint distribution in the present study, and assuming equal effects in patients on monotherapy and on combination therapy, we estimate ~92% power for a primary analysis in the PGx pazopanib population (at one-tailed α=0.05; N=1080), and ~61% power for a primary analysis in the PGx pazopanib monotherapy population (at one-tailed α=0.05; N=323). There are two key factors that may reduce the actual power: (i) The effect size estimate from the exploratory discovery study might be an overestimate because of winners' curse bias. (ii) The apparent effect size observed in the data analysed in the present study may be smaller because of heterogeneity among clinical studies, caused for example by systematic differences in patient characteristics, including (but not limited to) the presence of combination therapies. There is insufficient information to fully estimate the effect of these two factors on power.

To provide some robustness against the possibility that the genetic association effect size may be different in patients on combination therapy compared to patients on monotherapy, we considered a range of alpha spending strategies for the primary analysis, where we will declare significance if $P_M<\alpha_M$ in an analysis of patients on monotherapy alone and/or if $P_C<\alpha_C$ in a combined analysis of patients on monotherapy or on combination therapy, with $\alpha_M+\alpha_C=\alpha=0.05$ the overall false positive rate for the primary analysis. If effect sizes are equal in patients on monotherapy and on combination therapy, the optimal strategy is $\alpha_M=0$ and $\alpha_C=0.05$. On the other hand, if the effect size is zero in patients on combination therapy, the optimal strategy is $\alpha_M=0.05$ and $\alpha_C=0$. Although there is no exploratory data from which the relative effect sizes could be estimated (and hence the optimal strategy cannot be determined), numerical calculations suggest that values of α_C between 0.04 and 0.05 are close to optimal assuming an effect size in patients on combination therapy between 1x and 0.3x the effect size in patients on monotherapy (Figure 1). Therefore, for the primary analysis, $\alpha_M=0.01$ and $\alpha_C=0.04$ were chosen.

Figure 1 Power Estimate for *HLA-B57:01 with ALT Elevation**



For the *HLA-B**57:01 association, power and optimum alpha spend are plotted as functions of the effect attenuation on combination therapy, which is defined as the genetic effect size ratio between patients on combination therapy and patients on monotherapy. The three scenarios explore the impact of smaller true effect sizes relative to the point estimate from the exploratory study (scenario 1: same effect size; scenario 2: 25% smaller; scenario 3: 50% smaller). The solid lines show the optimal alpha spend on the combined analysis (α_C), assuming the remaining alpha (α_M=0.05−α_C) is spent on the monotherapy-only analysis. Dashed lines show power at the optimal alpha spend, and dotted lines show power at the chosen alpha spending strategy: α_M=0.01 and α_C=0.04.

4. GENETIC ANALYSIS POPULATIONS

Population	Definition / Criteria	Endpoint(s) Evaluated
PGx pazopanib	Comprises all subjects who receive at least one dose of pazopanib, as monotherapy or combination therapy for cancer, and who have given PGx consent and a sample and have been successfully genotyped This population will be based on the treatment the subject actually received.	All five endpoints

Population	Definition / Criteria	Endpoint(s) Evaluated
PGx pazopanib monotherapy	Comprises all subjects in the PGx pazopanib analysis population who were enrolled in trials or trial arms where the protocol did not specify treatment with specific combination therapy for cancer along with pazopanib	All five endpoints
PGx pazopanib combination therapy	Comprises all subjects in the PGx pazopanib analysis population who are not in the PGx pazopanib monotherapy analysis population	All five endpoints

5. CONSIDERATIONS FOR DATA ANALYSES

The genetic variants analysed will be as follows:

- *HLA-B**57:01, which will be imputed using the HIBAG algorithm (APPENDIX 10), using SNP genotyping data from the Affymetrix Axiom Biobank array
- Sixteen index SNPs with MAF>1% that achieved $P<5x10^{-7}$ in a previous exploratory genome-wide analysis:
    - rs62306729, rs80228453, rs148892667, rs113052844, rs148150732, rs187820820, rs148247629, rs13086084, rs6556844, rs114369408, rs12017140, rs151007454, rs139397837, rs9794884, rs17111888, rs215101

The primary analysis will evaluate carriage of a single allele (*HLA-B**57:01) for a single endpoint using a one-tailed test, and will have controlled false positive rate 5%.

Secondary analyses of association between *HLA-B**57:01 and other endpoints will be for effect size estimation and for exploratory purposes. Significant association with a secondary endpoint, but not with the primary endpoint, would not be considered a strict sense replication of the association observed in the exploratory analysis.

For secondary analyses of the 16 SNPs, false positives will be controlled at 5% for the primary endpoint (maximum on-treatment ALT), using a Bonferroni correction for 16 tests. Secondary analyses for these SNPs with other endpoints will be for effect size estimation and for exploratory purposes.

The PGx pazopanib monotherapy population will be analysed as part of a pre-specified subgroup analysis. To control the false positive rate, an alpha spending rule will be used for all primary analyses, spending $\alpha_M=0.01$ on the analysis in the PGx pazopanib monotherapy population and $\alpha_C=0.04$ on the analysis in the PGx pazopanib population (monotherapy and combination therapies combined).

6. DATA HANDLING CONVENTIONS

Table 2 provides an overview of appendices within this RAP for outlining data handling conventions.

Table 2 Overview of Appendices

Appendix	Component
1	Data Display Standards & Handling Conventions
2	Derived and Transformed Data
3	Premature Withdrawals & Handling of Missing Data
4	Genotype/Subject Quality Control

7. PHARMACOGENETIC ANALYSES

7.1. Primary Analyses

Primary Statistical Analysis
Endpoint / Covariates
• Maximum on-treatment ALT (in units of laboratory specific ULN), adjusted for clinical study and arm, ancestry PCs, sex, age at baseline, and baseline ALT (in ULN units). Normal linear regression will be used after appropriate transformation (log transform unless more aggressive transformation required).
Analysis Populations
• PGx pazopanib population • PGx pazopanib monotherapy population
Genetic Variants
• *HLA-B**57:01 using maximum weight imputed genotypes
Effects to be Modeled (Main or Interaction Effect)
• Dominant genetic model, coded 0/1 for absence or presence of *HLA-B**57:01
Statement Regarding What Constitutes a Significant Result
• P<0.04 for one-tailed test for carriage of *HLA-B**57:01 to be associated with higher on-treatment ALT in the PGx pazopanib population, and/or • P<0.01 for one-tailed test for carriage of *HLA-B**57:01 to be associated with higher on-treatment ALT in the PGx pazopanib monotherapy population

Sensitivity and Supportive Statistical Analysis
• Examination of effect within each clinical study and arm, by forest plot and tests of heterogeneity. Comparison of effect in PGx pazopanib monotherapy population vs. PGx pazopanib combination therapy population. • The impact of uncertainty about imputed *HLA* genotypes may be explored by using imputation

weights as genotype probabilities in a full likelihood based analysis (Kutalik, 2011)

- If more aggressive transformation is used for the primary analysis, then supportive analyses will include analyses of maximum on-treatment ALT using transformations that permit clinical interpretation (log or untransformed ALT in ULN units).

7.2. Secondary Analyses

Secondary Statistical Analysis
Endpoint / Covariates
Four endpoints representing different aspects of on-treatment ALT elevation: • Ever vs. never on-treatment ALT>5xULN • Ever vs. never on-treatment ALT>3xULN • Time until first on-treatment ALT>5xULN • Time until first on-treatment ALT>3xULN All analyses will be adjusted for clinical study and arm, ancestry PCs, sex, age at baseline, and baseline ALT (in ULN units). Logistic regression will be used for binary (ever vs never) endpoints, and Cox regression will be used for time until first event endpoints.
Analysis Population
• PGx pazopanib analysis population • PGx pazopanib monotherapy population
Genetic Variants
• *HLA-B**57:01
Effects to be Modeled (Main or Interaction Effect)
• Dominant genetic model, coded 0/1 for absence or presence of *HLA-B**57:01
Statement Regarding What Constitutes a Significant Result
For each endpoint: • P<0.04 for one-tailed test for carriage of *HLA-B**57:01 to be associated with higher on-treatment ALT in the PGx pazopanib population, and/or • P<0.01 for one-tailed test for carriage of *HLA-B**57:01 to be associated with higher on-treatment ALT in the PGx pazopanib monotherapy population Since no multiple testing adjustment is made for analyzing multiple endpoints, then in the absence of significant association in the primary analysis, significant associations with these endpoints would be regarded as exploratory results and not as a strict sense replication of the association observed in the exploratory analysis.

Secondary Statistical Analysis
Endpoint / Covariates
• Maximum on-treatment ALT (scaled by ULN) • Ever vs never on-treatment ALT>5xULN

• Ever vs never on-treatment ALT>3xULN • Time until first on-treatment ALT>5xULN • Time until first on-treatment ALT>3xULN All analyses will be adjusted for clinical study and arm, ancestry PCs, sex, age at baseline, and baseline ALT (in ULN units). Normal linear regression will be used for maximum on-treatment ALT, after suitable transformation, as for the primary analysis. Logistic regression will be used for binary (ever vs never) endpoints, and Cox regression will be used for time until first event endpoints.
Analysis Population
• PGx pazopanib analysis population • PGx pazopanib monotherapy population
Genetic Variants
A total of 16 index SNPs identified in previous analyses.
Effects to be Modeled (Main or Interaction Effect)
• Effect of genetic variant, coded 0/1/2 for copies of the non-reference allele (additive genetic model). For SNPs not directly genotyped, imputed dosage of the non-reference allele will be used.
Statement Regarding What Constitutes a Significant Result
• P<0.04/16 for one tailed test for each genetic variant, for the primary endpoint (maximum on-treatment ALT) in the PGx pazopanib population, and/or • P<0.01/16 for one tailed test for each genetic variant, for the primary endpoint (maximum on-treatment ALT) in the PGx pazopanib monotherapy population. The criteria above control the FWER at 5% for this secondary analysis. The same significance thresholds will be used for other endpoints, but since no further multiple testing adjustment is made for analyzing multiple endpoints, significant associations (only) with these other endpoints would be regarded as exploratory results.

7.3. Other Exploratory Analyses

7.4. General Pharmacogenetic Analysis Conventions

Table 3 provides an overview of appendices within the RAP for outlining general pharmacogenetic analysis conventions.

Table 3 Overview of Appendices

Appendix	Component
5	Multiple Comparisons and Multiplicity
6	Hardy-Weinberg Equilibrium (HWE) Analysis
7	Linkage Disequilibrium Analysis
8	Characterizing Ancestry Using Principal Components Analysis
9	Estimation of Heritability
10	Genotype Imputation

8. REFERENCES

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics.* 2012;44:955-9.

Kutalik Z, Johnson T, Bochud M, Mooser V, Vollenweider P, Waeber G, Waterworth D, Beckmann JS, Bergmann S. Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics.* 2011;12:1-17.

Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual Review of Genomics and Human Genetics.* 2009;10:387-406

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26(22):2867-73.

Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol.* 2008;32:567.

Novembre J, Johnson T, Bryc K, Kuralik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. *Nature*. 2008; 456:98-101.

Patterson V, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2(12): e190

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;**38**(8):904-909. PMID 16862161.

The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1016-1073. PMID 20981092.

Vattikuti S, Guo J, Chow CC. Heritability and Genetic Correlations Explained by Common SNPs for Metabolic Syndrome Traits. PLoS Genet 2012;8(3): e1002637. doi:10.1371/journal.pgen.1002637.

Xu CF, Reck BH, Goodman VL, Xue Z, Huang L, Barnes MR, Koshy B, Spraggs CF, Mooser VE, Cardon LR, Pandite LN. Association of the hemochromatosis gene with pazopanib-induced transaminase elevation in renal cell carcinoma. *J. Hepatology.* 2011;54:1237-43.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42:565-9.

Zheng X, Shen J, Cox C, Wakefield J, Ehm M, Nelson M, Weir BS. HIBAG -- HLA genotype imputation with attribute bagging. *Pharmacogenomics Journal*. 2014;14:192-200.

Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL *et al*. Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. *Nat. Genet.* 2011; 43:117–20.

9. APPENDICIES

Appendix Number	Appendix Description
Gx RAP Section 6 : Data Handling Conventions	
APPENDIX 1	Data Display Standards & Handling Conventions
APPENDIX 2	Derived and Transformed Data
APPENDIX 3	Premature Withdrawals & Handling of Missing Data
APPENDIX 4	Genotype/Subject Quality Control
Gx RAP Section 6 : General Genetic Analysis Conventions	
APPENDIX 5	Multiple Comparisons and Multiplicity
APPENDIX 6	Hardy-Weinberg Equilibrium (HWE) Analysis
APPENDIX 7	Linkage Disequilibrium Analysis
APPENDIX 8	Characterizing Ancestry Using Principal Components Analysis
APPENDIX 9	Estimation of Heritability
APPENDIX 10	Genotype Imputation
Other Gx RAP Appendices	
APPENDIX 11	Abbreviations & Trade Marks

9.1. APPENDIX 1: Data Display Standards & Handling Conventions

The number of patients included in each analysis population will be summarized by endpoints and baseline characteristics. In general, categorical data will be summarized using frequency counts and percents, and continuous data will be summarized using means, standard deviations, percentiles (e.g. minimum, 1st quartile, median 3rd quartile and maximum). Summaries will be calculated for each analysis population overall, and if appropriate in relevant subgroups.

Genetic associations will be summarized by regression model effect size estimates and standard errors, adjusted for covariates. Effect size estimates and confidence interval endpoints may be transformed from the analysis scale (such as log odds ratio or log hazard ratio) to an alternative scale to facilitate interpretation (such as odds ratio or hazard ratio). P-values will be calculated using an F test for normal linear models and using a likelihood ratio test for generalized linear models and Cox regression. Associations may be displayed using an appropriate plot or table of endpoint versus genotype (such as dotplot or boxplot for continuous endpoints, Kaplan—Meier estimates of survival or cumulative incidence function, contingency table for binary or categorical endpoints). Genotype or endpoint categories may be combined to generate 2x2 contingency tables when calculation of genotype test sensitivity, specificity, positive or negative predictive value may facilitate interpretation.

9.2. APPENDIX 2: Derived and Transformed Data

Should the distribution of any dependent variable deviate substantially from that assumed for a particular analysis method, an appropriate transformation will be applied or a robust method used.

9.3. APPENDIX 3: Premature Withdrawals & Handling of Missing Data

9.3.1. Premature Withdrawals

Patients who withdrew consent for the optional PGx research component of the clinical studies prior to genetic consent reconciliation for this PGx study are not included in this analysis.

9.3.2. Handling of Missing Genetic Data

The endpoint, covariates, key demographic/baseline variables and time on study may be compared between the Genetic analysis population against individuals not analyzed for PGx. The summary statistics will be inspected for any concerning imbalances. If any imbalances that may affect the analysis are identified, these factors may be explored further and/or accounted for in the analysis models.

9.4. APPENDIX 4: Genotype/Subject Quality Control

9.4.1. Subject Quality Control

Subjects will be excluded according to the following criteria: (i) subjects with arrays where genotyping failed, as identified in the manufacturer's genotype calling software and following manufacturer's guidelines; (ii) subjects with low call rate (threshold to be determined based on the data); (iii) subjects for whom sex inferred from sex chromosome genotypes cannot be reconciled with sex recorded on the CRF (e.g. sample swap); (iv) subjects with identical genotypes (e.g. identical twins, multiple participation for same individual or sample plating errors); (v) subjects with high-degree of cryptic relatedness. Following subject exclusions and before the statistical analysis, SNP exclusions will be applied as part of genotype imputation as described in Section 9.4.2.

Cryptic relatedness refers to a situation where multiple individuals in a study sample are genetically related to one another, which if present to a substantial degree could bias analysis results. A software tool, KING [Manichaikul, 2010], will be used to check family relationship by estimating all kinship coefficients for all pairwise relationships. For pairs of DNA samples that have 3rd-degree relationship or more closer, one sample in each pair will be excluded from the analysis.

9.4.2. Genotype Quality Control

Prior to genotype imputation (see APPENDIX 10), variants in each GWAS dataset will be excluded if they have low call rate, if they have poor calling metrics, if they show deviations from Hardy-Weinberg proportions within subgroups of any given ancestry (see APPENDIX 6), if they are monomorphic, if they show gross and irreconcilable differences in alleles or allele frequency with reference panel genotypes from the HapMap or 1000 Genome projects. After imputation, QC metrics will be examined to identify strand flip errors (e.g. correlation between measured and imputed genotype close to r=-1) and if necessary these variants will be removed and imputation rerun. Post-imputation, there will be no missing genotype data. Variants will not be excluded post-imputation on the basis of minor allele frequency/count or imputation quality metrics, unless inspection of association statistic QQ and Manhattan plots suggests excess false positive associations [Kutalik, 2011].

9.5. APPENDIX 5: Multiple Comparisons & Multiplicity

9.5.1. Candidate variant analysis

Bonferroni corrections and an a priori chosen alpha spending rule will be used, as described in the main body of this RAP.

9.6. APPENDIX 6: Hardy-Weinberg (HW) Analysis

Hardy-Weinberg (HW) proportions is a historic term for the notion that alleles are inherited from each parent independently, and thus expected genotype frequencies can be predicted from allele frequencies. Departure from HW proportions can have several causes, including genotyping error, and admixture of subjects with different ancestries. HW analysis will be conducted for all genotyped variants and will be conducted within race and ethnicity groups that have sufficient sample sizes. For variants significantly associated with any endpoint, substantial evidence of departure from HW proportions will be investigated for possibility of genotyping error (e.g. by manual examination of cluster plots, and by examination of variants that should be in linkage disequilibrium with the focal variant).

9.7. APPENDIX 7: Linkage Disequilibrium Analysis

Linkage Disequilibrium (LD) measures the association between alleles at different loci. It can help understand if association signals in the same region are independent from each other or due to correlation among the variants. LD analysis may be conducted for interesting variants, if appropriate, using subjects from the population of interest. Pairwise LD will be limited to variants located within a particular gene or gene region

9.8. APPENDIX 8: Characterizing Ancestry Using Principal Components Analysis

Principal component analysis (PCA) of large numbers of genetic variants (typically genome-wide) can be used to characterize ancestry for each genotyped subject [Price, et al. 2006, Patterson, et al. 2006 , Novembre, et al. 2008]. The principal components may be used as covariates in tests of genetic association (e.g. regression of an endpoint onto each individual genetic variant in turn), to correct for confounding due to population stratification [Price, 2006]. Clustering based on the principal components may also be used to refine self-reported race and ethnicity to facilitate investigation of genetic effects specific to certain ancestry groups.

9.9. APPENDIX 9: Estimation of Heritability

A mixed model method developed by [Yang, *et al*. 2010] may be used to estimate the combined contribution of all common variants to the heritability of interesting variables. This approach, which makes use of genome wide variants in sets of presumably unrelated

individuals, has been used to estimate the heritability of measured variants to numerous human traits and diseases, including height [Yang, *et al.* 2010], quantitative traits associated with metabolic syndrome [Vattikuti, *et al*. 2012], and response to treatment of type 2 diabetes with metformin [Zhou, *et al.* 2011]. The software package developed by Visscher and colleagues, GCTA, will be used to carry out the estimation procedure.

9.10. APPENDIX 10: Genotype Imputation

Genotype imputation for genetic variants that were not genotyped on the Axiom Biobank array ("untyped variants") will be performed using a cosmopolitan haplotype reference panel from the 1000 Genomes Project, and using Hidden Markov Model methods as implemented in MaCH and minimac [Li, 2009] [Howie, 2012]. APPENDIX 4 describes subject and SNP exclusions that will be applied prior to imputation.

HLA genotype imputation will be performed using the HIBAG algorithm and published parameter estimates [Zheng, 2014].

9.11. APPENDIX 11: Abbreviations & Trade Marks

9.11.1. Abbreviations

Abbreviation	Description
ALT	Alanine aminotransferase
CTCAE	Common Terminology Criteria for Adverse Events
DNA	Deoxyribonucleic acid
CRF	Case Report Form
Gx RAP	Genetics Reporting & Analysis Plan
GSK	GlaxoSmithKline
HWE	Hardy-Weinberg Equilibrium
LD	Linkage Disequilibrium
PC	Principal Component
PCA	Potential Component Analysis
PGx	Pharmacogenetics
RAP	Reporting & Analysis Plan
SNP	Single Nucleotide Polymorphism
ULN	Upper Limit of Normal

9.11.2. Trademarks

Trademarks of the GlaxoSmithKline Group of Companies
VOTRIENT

Trademarks not owned by the GlaxoSmithKline Group of Companies
None