Division: Worldwide Development

Information Type: Pharmacogenetics Reporting and Analysis Plan

Title:	Pharmacogenetics Reporting and Analysis Plan for Genetic Evaluation of Pazopanib–Related Hepatotoxicity (PGx6652, VEG117365)
Compound Number	GW786034
Effective Date:	19-MAR-2013

Brief Description:

This document details the reporting and analysis plan for VEG117365 (PGx6652), genetic evaluation of pazopanib-related hepatotoxicity in subjects from multiple clinical studies. The objective of this exploratory pharmacogenetic analysis is to identify genetic markers associated with pazopanib-related hepatotoxicity. Both a candidate gene analysis and a genome wide association study (GWAS) analysis will be conducted. The candidate gene analysis will evaluate association between candidate genetic markers and serious liver injury and will include data from pazopanib-treated subjects with concurrent elevation of alanine aminotransferase (ALT) and total bilirubin from 16 clinical studies and a set of race and ethnicity matched controls. The GWAS analysis will test association between genetic variants and hepatotoxicity and will include data from eight Phase II or III pazopanib clinical studies (VEG102616, VEG105192, VEG107769, VEG108844, VEG113078, VEG110727, VEG110655, and VEG114012).

Subject: Pharmacogenetics (PGx), genome wide association study (GWAS), human leukocyte antigen (HLA), renal cell carcinoma (RCC), soft tissue sarcoma, ovarian cancer , pazopanib (GW786034), hepatotoxicity, alanine aminotransferase (ALT), total bilirubin (TBL).

Author's Name(s), Title and Affiliation:

[redacted] Scientific Investigator, Statistical Genetics, Quantitative

[redacted] Senior Scientific Investigator, Statistical Genetics, Quantitative Sciences

[redacted] Director, Genetics, Quantitative Sciences

Copyright 2013 the GlaxoSmithKline group of companies. All rights reserved. Unauthorized copying or use of this information is prohibited.

Approved by:

[redacted]	**March 19th , 2013**
Director, Statistical Genetics, Quantitative Science	**Date**

Distribution List:

[redacted]	Director, Clinical Development, Oncology
[redacted]	Principal Scientist, Genetic Data Sciences, Quantitative Sciences
[redacted]	Director, Clinical Development, Oncology
[redacted]	Director, Clinical Development, Oncology
[redacted]	Director, Clinical Development, Oncology
[redacted]	Manager, Statistics & Programming, Quantitative Sciences
[redacted]	Director, Statistics & Programming, Quantitative Sciences
[redacted]	Investigator, Genetics, Quantitative Sciences

TABLE OF CONTENTS

PAGE

ABBREVIATIONS & GLOSSARY....4

1. INTRODUCTION....6

2. STUDY OBJECTIVES AND ENDPOINTS....6
   - 2.1. Study Objectives....6
   - 2.2. Study Endpoints....7

3. STUDY DESIGN....7

4. GENERAL CONSIDERATIONS FOR DATA ANALYSES....8
   - 4.1. Power Calculations....8
   - 4.2. Multiple testing corrections....10

5. ANALYSIS PLAN....10
   - 5.1. Analysis populations....10
   - 5.2. Dependent variables....10
   - 5.3. Genotype variable definitions....11
   - 5.4. Data Quality Control....11
     - 5.4.1. Subject and Genotype data quality control....11
     - 5.4.2. Genotype Imputation....12
     - 5.4.3. Control for population stratification and ancestry....12
     - 5.4.4. Control for other independent variables (covariates)....13
   - 5.5. Statistical analysis....13

6. DATA HANDLING CONVENTIONS....15
   - 6.1. Premature Withdrawal and Missing Data....15

7. REPORTING CONVENTIONS....15
   - 7.1. Demographic, Baseline and Endpoint Characteristics....15
   - 7.2. Summary of Genetic Association Analysis Results....15

8. REFERENCES....17

9. APPENDIX....20

ABBREVIATIONS & GLOSSARY

Allele	Any one of several alternative forms of a gene or DNA sequence at the same genetic locus
ALT	Alanine aminotransferase
DNA	Deoxyribonucleic acid
FWER	Family-wise error rate
Haplotype	The alleles in a given subject that were inherited from one biological parent
Genetic locus (plural loci)	A region at an identifiable position on a given chromosome, which can be an entire gene, part of a gene, or a segment of DNA of unknown function
Genetic variant	A genetic locus at which two or more distinct alleles have been observed
Genotype	The alleles present in a given subject. At each autosomal genetic locus, one allele is inherited from each biological parent
GWAS	Genome-Wide Association Study. A study in which alleles at genetic loci throughout the genome are tested for association, in a way that is not biased in favour of genetic loci in or near previously hypothesized candidate genes
HLA	Human leukocytic antigen
HWE	Hardy Weinberg Equilibrium (HWE) analysis provides a measure of the association between the two alleles present in the same subject at a given genetic locus. Genotyping error is a common cause (although not the only cause) of departure from HWE
ITT	Intent to treat
OR	Odds ratio
PC	Principal component
PCA	Principal components analysis
PGx	Pharmacogenetic
PPV	Positive predictive value
QC	Quality control
QQ plot	Quantile-Quantile (QQ) probability plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other
RAF	Risk allele frequency
RAP	Reporting and analysis plan
RCC	Renal cell carcinoma
SE	Standard error
SNP	Single nucleotide polymorphism
TBL	Total bilirubin
ULN	Upper limit of normal
VEGFR	Vascular endothelial growth factor receptors

Trademark Information

Trademarks of the GlaxoSmithKline group of companies
VOTRIENT™

Trademarks not owned by the GlaxoSmithKline group of companies
None

1. INTRODUCTION

Pazopanib is an oral angiogenesis inhibitor targeting vascular endothelial growth factor receptors (VEGFR) -1, -2, and -3, platelet-derived growth factor receptors -α, and -β, and the stem cell factor receptor, c-Kit. It has been approved for the treatment of advanced renal cell carcinoma (RCC) and soft tissue sarcoma, and is in clinical development for other indications. Elevations in serum alanine aminotransferase (ALT) and serum total bilirubin (TBL) were observed in pazopanib clinical studies [Sternberg, 2010] [van der Graaf, 2012]. For instance, ALT ≥3 times the upper limit of normal (ULN) occurred in 18% and 3% of the subjects with RCC who received pazopanib and placebo, respectively [Sternberg, 2010]. Concurrent elevations of ALT and bilirubin without alkaline phosphatase elevation, conditions which may be associated with severe liver injury, were rare (~1%). Previous pharmacogenetic (PGx) studies in pazopanib-treated subjects with advanced RCC using data from clinical studies VEG102616 and VEG105192 have found that the Gilbert's syndrome associated *UGT1A1* variant was associated with pazopanib-related bilirubin elevation [Xu, 2010]. Although a genetic association between *HFE* polymorphisms and ALT was also observed [Xu, 2011], this observation was limited by small sample size and would require further validation.

The purpose of this exploratory PGx analysis is to evaluate associations between genetic variants and hepatotoxicity in pazopanib-treated subjects with cancer. This PGx analysis is based on the assumption that the genetic basis for the risk of pazopanib-related hepatotoxicity is independent of cancer type and will use data from multiple pazopanib clinical studies [e.g. RCC, soft tissue sarcoma, and ovarian cancer (Appendix Table 1 and Appendix Table 2)].

Human leukocytic antigen (HLA) polymorphisms within the major histocompatibility complex are strongly associated with hepatotoxicity for a number of drugs, including amoxicillin-clavulanate (*HLA-DRB1*15:01*), anti-tuberculosis chemotherapy (*HLA-DQB1*02:01* and *HLA-A*03:01*), ticlopidine (*HLA-A*33:03*), ximelagatran (*HLA-DRB1*07:01*), flucloxacillin (*HLA-B*57:01*), lumiracoxib (*HLADRB1*15:01*), and lapatinib (*HLA-DRB1*07:01*) [Harper, 2012] [Spraggs, 2012]. Furthermore, pazopanib trough plasma concentration may impact subjects' risk for elevated ALT [Pandite, 2010]. Therefore, association between genetic variants in *HLA* and in genes involved in pazopanib metabolism and disposition and the risk of liver injury will be evaluated in a candidate gene analysis.

In addition, a genome wide association study (GWAS) analysis will be conducted to explore genetic associations with ALT elevation, an important indicator for liver injury.

2. STUDY OBJECTIVES AND ENDPOINTS

2.1. Study Objectives

The objective is to identify genetic markers associated with pazopanib-related hepatotoxicity, by applying the following two approaches:

- Evaluate associations between candidate genetic markers and serious liver injury in pazopanib-treated subjects (referred to as "**candidate gene analysis**" in subsequent text)
- Test associations between genetic variants and hepatotoxicity in pazopanib- treated subjects using a genome wide association study approach (referred to as "**GWAS analysis**" in subsequent text)

2.2. Study Endpoints

In the candidate gene analysis, serious liver injury cases are defined as pazopanib-treated subjects with cancer who experienced concurrent ALT≥3x ULN and total bilirubin ≥2xULN, excluding subjects with obstructive jaundice or biliary obstruction. Controls who satisfy ALL of the following five criteria are selected from three RCC studies (see Section 3): a) at least 12 weeks exposure to pazopanib at 800mg/day; b) having no more than two missing records of ALT and total bilirubin measurements for scheduled visits in the first 12 weeks of the on-therapy window; c) both ALT and total bilirubin≤1xULN for all non-missing records in the on-therapy window; d) available genotyping data for the candidate markers; and e) can be matched to at least one case individual by self-reported race and ethnicity.

In the GWAS analysis, two primary endpoints will be derived from serum ALT levels for each subject:

- Maximum ALT measured within the on-therapy window;
- Time-to-event defined as the time from initiation of pazopanib treatment until the first on-therapy event (ALT measure ≥3xULN). Subjects with no on-therapy measure of ALT>ULN will be censored at the end of the on-therapy window. Subjects with maximum on-therapy ALT measure >ULN but <3xULN will be excluded from this time to event analysis.

Additional exploratory analyses may be conducted to explore genetic associations by incorporating more clinical/lab data, such as the full time series of ALT, AST and serum total bilirubin measures, and exposure to treatment data. In addition, pending review of the initial analysis results, events may be defined using alternative ALT thresholds (e.g.5xULN, 10xULN) or for having concurrent ALT and bilirubin elevations.

3. STUDY DESIGN

In the candidate gene analysis, associations between HLA alleles and functional markers in genes involved in pazopanib metabolism and disposition (Appendix Table 3) and the pre-defined case/control status will be tested. Additional candidate gene markers maybe added pending emerging literature information or after review of the initial analysis results. All subjects included in the analysis were selected from prospective clinical trial cohorts. For notational convenience, the subjects in this analysis will be referred as "cases" and "controls" using the criteria described in Section 2.2. In total, there will be ~32 cases and ~68 controls in the analysis. Genotype data for the serious liver injury cases who experienced concurrent ALT and bilirubin elevation from 16 clinical studies (Appendix Table 1) will be generated by: 1) sequencing for HLA alleles;

2) Affymetrix DMET array for functional genetic markers in genes involved pazopanib metabolism and disposition; and 3) sequencing or IVD assay for *UGT1A1**28. Controls will be taken from three clinical studies (VEG102616, VEG105192 and VEG107769) where genotyping data for the candidate markers are already available from previous analysis [Xu, 2011].

In GWAS, association analyses will be conducted using data from ~1273 pazopanib-treated subjects from eight phase II or III clinical studies (VEG102616, VEG105192, VEG107769, VEG110727, VEG108844, VEG113078, VEG110655 and VEG114012) (Appendix Table 2). Genome wide SNP genotyping has been conducted using three different arrays manufactured by Illumina (San Diego, CA): Human1M BeadChip for VEG102616, VEG105192 and VEG107769 (N=403); HumanOmni5-quad beadarray for VEG110727, VEG108844 and VEG113078 (N=471); and HumanOmniExpressExome for VEG110655 and VEG114012 (N=399). Approximately 50% of SNPs on the Human1M are genotyped on both of the other two arrays. Untyped SNP genotypes will be imputed, using haplotype reference panels from the 1000 Genomes (1000G) Project. Additional targeted genotyping has been generated to ensure availability of data for functional genetic variants in genes involved in pazopanib mode of action or metabolism and disposition (Appendix Table 4). Specific hypotheses of association with ALT elevation will be implemented as “candidate gene lookup” analyses within GWAS, where P-values from the genome wide association analysis results will be reported, regardless of whether they reach genome-wide significance, for specific genetic variants from genes with biological plausibility to the endpoint variables. For inclusion in the GWAS analysis, subjects must have a baseline ALT measurement and at least one on-therapy ALT measurement.

4. GENERAL CONSIDERATIONS FOR DATA ANALYSES

4.1. Power Calculations

In the candidate gene analysis, with assumption of 1% population prevalence of concurrent ALT and bilirubin elevation in pazopanib-treated subjects [Goodman, 2010], type-I error α =0.05 after multiple comparison correction based on 372 independent tests and sample size of 32 cases and 68 controls, statistical power in excess of 60% is achieved for plausible genetic models with homozygous odds ratio (OR) in 5-20 range and risk allele frequency (RAF) between 10-50% (Figure 1). Greater than 80% statistical power could be achieved for all genetic models that have both high positive predictive value (PPV) and high sensitivity (Figure 2).

Figure 1 Power of Case-Control Candidate Gene Analysis

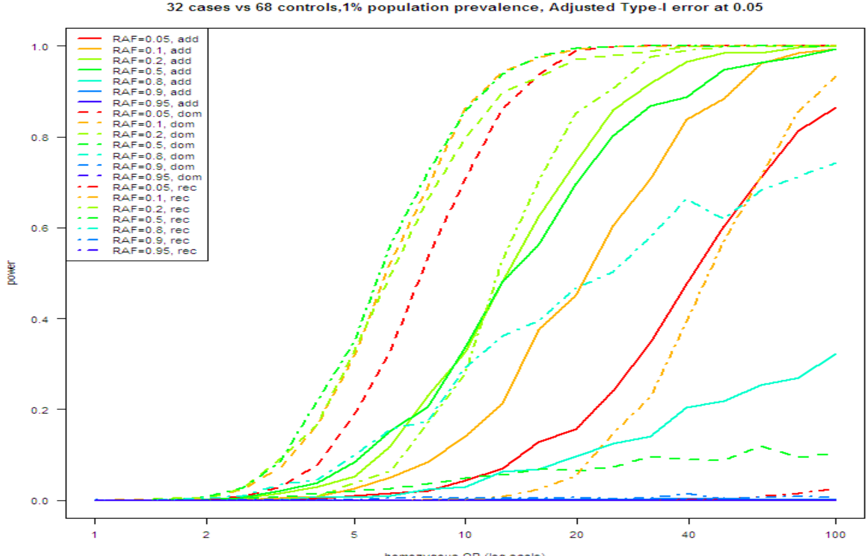
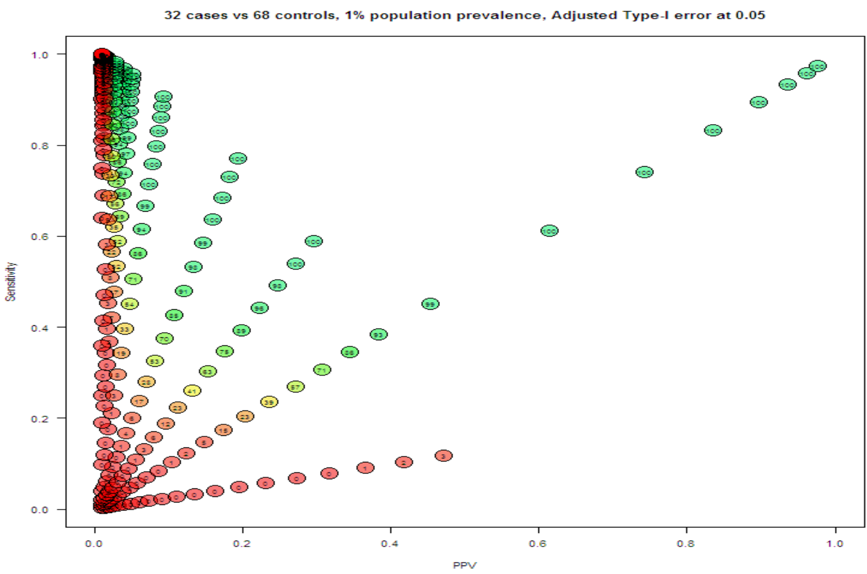


Figure 2 Power of Candidate Gene Analysis as Function of Sensitivity and PPV *



* Power was written in numerals as a percentage and colored red/green for low/high power.

In **GWAS**, greater than 80% statistical power to reach genome-wide significance ($P \leq 5x10^{-8}$) is achieved in analysis of data from 1273 individuals, for genetic variants that explain 3.1% or more phenotypic variance in a quantitative trait analysis. Assuming a phenotypic standard deviation of 0.42 for log10(max ALT), then for common genetic variants (minor allele frequency between 0.1 and 0.5), this corresponds to effect sizes between 1.27 and 1.50 fold difference in max ALT per allele under an additive genetic model.

4.2. Multiple testing corrections

For the candidate gene analysis, permutation tests will be used to control type I (false positive) error under multiple testing. Permutations will permute the relationship between genetic marker and clinical endpoint variable within race/ethnicity strata. This generates the empirical distribution of association statistics under the assumed null hypothesis. The empirical P-value, which will have controlled for family wise error rate (FWER) at 5%, will be calculated for each tested genetic marker.

For the GWAS analysis, the conventional $P \leq 5x10^{-8}$ threshold for declaring genome-wide significance will be used [McCarthy, 2008] [Dudbridge, 2008]. If none of the genetic markers achieve genome-wide significance, top-hit SNPs based on either their ranks (e.g. top 100 most significant SNPs) or applying less stringent P-value thresholds (e.g. SNPs with $P \leq 5x10^{-5}$) may be followed up. These ad-hoc criteria will not be determined until all analyses are done and the results are seen. For the "candidate gene lookup" analyses within the GWAS, a Bonferroni correction for the number of associations looked up will be used (Appendix Table 4).

Gene-enriched and pathway-based analyses may be explored. These approaches limit the total number of association tests to gene or pathway level, which is significantly reduced comparing to the number of tests for all SNPs genome-wide. The scope of this optional analysis will be determined after the main analyses are accomplished.

5. ANALYSIS PLAN

5.1. Analysis populations

Candidate gene analysis: The analysis population will consist of subjects who: (i) provided written informed consent for genetics research; (ii) provided sufficient DNA sample for genotyping and were successfully genotyped for at least one candidate genetic marker; (iii) received at least one dose of pazopanib; (iv) satisfied the case or control definitions.

GWAS analysis: The analysis population will consist of subjects who: (i) provided written informed consent for genetics research; (ii) provided sufficient DNA sample for genotyping and were successfully genotyped on an array; (iii) received at least one dose of pazopanib; (iv) have a baseline ALT measure and at least one ALT measure in the on-therapy window.

5.2. Dependent variables

Binary form (case vs. control) dependent variables will be used in the **candidate gene analysis**; two continuous endpoints will be derived from serum ALT levels and used in the **GWAS analysis**. See Section 2.2 for definition of case-control status and derivation of continuous endpoints.

Different forms of dependent variables may be explored in additional exploratory analyses which take into account factors such as the variation in exposure to treatment (e.g. dose, duration, or cumulative dose). In addition, multivariate analysis approaches

may be used, including the creation of further derived dependent variables (composite phenotypes) that incorporate ALT measures at multiple time points.

5.3. Genotype variable definitions

Appendix Table 3 summarizes information about genes and markers to be included in the **candidate gene analysis**.

In the **GWAS analysis**, genetic variants will be mapped and alleles will be coded for the + strand on a suitable build of the human reference sequence (currently hg19; GRC build 37).

5.4. Data Quality Control

5.4.1. Subject and Genotype data quality control

For the candidate gene analysis, subjects will be removed from this analysis if subjects fail genotyping for all candidate genetic markers or the genotypes for their sex chromosomes do not agree with their reported gender (pending data availability). HLA data QC was conducted by vendors prior delivery of data based on their SOP. For Affy DMET genetic marker quality control, genotype data will be considered valid if the call rate on the chip is >95%.

For the GWAS analysis, subjects will be excluded according to the following criteria: (i) subjects with arrays where genotyping failed, as identified in the manufacturers 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 gender inferred from sex chromosome genotypes cannot be reconciled with gender recorded on the CRF (e.g. sample swap); (iv) subjects with identified identical genotypes (e.g. identical twins, multiple participation for same individual or sample plating errors); (v) subjects with high-degree of relatedness. Following subject exclusions and before the statistical analysis, SNP exclusions will be applied as part of genotype imputation as described in Section 5.4.2.

Because genotyping performance varies for different arrays, the exclusion thresholds will be chosen after review of summary QC metrics of the genotyping data. The proposed procedure may be adapted according to different scenarios that may arise. For example, when cryptic relationships among study participants are detected, the solution usually depends on the number and extent of detected issues. If the incidences are very rare with relationship beyond second degree relatives, the issue can be ignored during analysis; if the incidences are rare but the relationship is between closely linked family members (e.g. parent-child, sibs etc.), subject with lower genotyping rate from each related pair will be excluded; if the incidence of high pairwise relatedness is common, modified association tests will be used that correct for relatedness using a mixed model approach with the correlation matrix among study subjects will be considered [Zhou, 2012].

5.4.2. Genotype Imputation

Because subjects from different clinical studies were genotyped using different genome wide SNP arrays (see Appendix Table 2), genotype imputation will be performed, so that association statistics can be meta-analysed across clinical studies in a simple way, for each genetic variant in turn. This analysis strategy has been widely successful in complex trait GWAS analyses [Kolz, 2009] [Teslovich, 2010]. The imputation of genotypes at genetic variants that were not genotyped on each genome wide SNP array (hereafter "untyped variants") will be performed using cosmopolitan haplotype reference panels from the 1000G Project, and using hidden Markov model methods as implemented in MaCH and minimac [Li, 2009] [Howie, 2012]. Genotype imputation may be performed separately for each clinical study, or for groups of clinical studies for the same indication that were genotyped using the same platform (e.g. VEG102616, VEG105192 and VEG107769). In addition, additional exploratory analysis may use HLA genotypes imputed for all subjects with genome wide SNP data. HLA genotype imputation will be performed using the HIBAG algorithm and published parameter estimates (available R-package from (http://cran.r-project.org/web/packages/HIBAG/index.html).

Prior to genotype imputation, markers in each GWAS dataset will be excluded if they have low call rate (say <98%), if they have poor calling metrics, if they show deviations from Hardy-Weinberg Equilibrium (HWE, say $P<10^{-7}$ analysed within ancestry groups), if they are monomorphic, if they show gross and irreconcilable differences in alleles or allele frequency with reference panel genotypes from the HapMap or 1000G 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 markers will be removed and imputation rerun. Post-imputation, there will be no missing genotype data. Markers 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].

5.4.3. Control for population stratification and ancestry

In the candidate gene analysis, the matched case-control study design is expected to reduce the possibility of confounding due to population stratification. As caveats, even though unlikely, it is still possible that confounding factors exist due to (i) discrepancies between self-reported race/ethnicity and genetic ancestry; (ii) potential stratification within certain population (e.g. Eastern vs. Western European). Hence, future studies will be required to replicate any associations identified in this study.

In the GWAS analysis, ancestry principal components analysis (PCA) will be used to cluster subjects [Price, 2006] [Patterson, 2006] [Novembre, 2008]. Clustered subjects will be plotted in reference to individuals in HapMap control panel with confirmed ancestry. Self-reported race/ethnicity for each subject will be compared against their genetic defined ancestry subgroup. Subjects who are outliers in the ancestry PCA will be flagged for exclusion in sensitivity analyses. PCA may also be used to detect potential stratification of the analysis population due to batch effect produced in sample preparation and genotyping. Finally, these identified confounding factors will be controlled through further quality control measures or during statistical analysis (e.g.,

including the first few PCs representing major factors as covariates in regression analysis models).

5.4.4. Control for other independent variables (covariates)

Independent genetic variables are defined in Section 5.3. Additional independent variables necessary to control for population structure and race/ethnicity are defined in Section 5.4.3.

In the **primary candidate gene analysis**, no further independent variables will be adjusted in the analysis model except for race and ethnicity. Inclusion of additional covariates (e.g. baseline ALT, concomitant medication, tumour type, pazopanib exposure, age, gender, etc) may be considered in additional exploratory analyses.

In the **primary GWAS analysis**, two regression models will be used: one will include only PCs as covariates to control for possible confounding effects and the other will include age, sex, baseline ALT, and PCs as covariates. Additional covariates may be included in the analysis model in further exploratory analysis after reviewing the demographic, laboratory and clinical data (e.g. BMI, cancer type, pazopanib exposure, concomitant medications); this will be determined upon group discussion. Selection criteria will be based on their biological and the statistical relevance to the interested endpoint variables.

5.5. Statistical analysis

For the **candidate gene analysis**, association between candidate gene alleles/genotypes and the case/control endpoint will be tested using the logistic regression analysis with appropriate adjustment of covariates. For the **GWAS analysis**, association between the continuous form of endpoints and genetic variants across the entire genome will be tested using linear and Cox regressions. Details of the analysis methods are described below. In both **candidate gene and GWAS analysis**, additive genetic model will be used initially. Other genetic models/tests (e.g. dominant/recessive models or 2d.f. genotype test) will be explored after reviewing of the initial analysis results.

For the candidate gene analysis, for each HLA gene, association will be tested by contrasting each observed allele against all others combined. For example, for HLA-A*01:01, the allele will be coded as either X or 0101, where X represents all alleles other than 0101. For genetic variants with two alleles including the recoded HLA genes, the genotype will be codes as integers of 0/1/2 for an additive genetic model or 0/1 for a dominant or recessive genetic model. Specifically for genes involved in pazopanib metabolism and disposition, alleles will be pooled based on a priori defined functional categories, e.g. poor /normal/extended metabolizer will be coded 0/1/2 respectively. In the primary analysis of genetic association, only race and ethnicity will be included as covariates.

Upon completion of primary analyses, additional exploratory analyses may be conducted such as: (i) adjusting for additional covariates (Section 5.4.4); (ii) haplotype analysis to study multiple alleles simultaneously within the same gene or region; (iii) incorporating multiple HLA genes in the regression model and test how they act together to affect the

variation of dependent variables [Yang, 2012]; (iv) convert HLA 4 digit alleles to their corresponding amino acid sequences, and test for association between each amino acid change and case/control status [Raychaudhuri, 2012].

For the GWAS analysis, clinical and genotype data from the eight clinical studies will be analyzed jointly. Maximum on-treatment ALT level or time-to- the first ALT ≥3xULN event will be treated as continuous dependent variables. The endpoints will be regressed onto coded genotypes, adjusting for appropriate sets of covariates. In the first step, GWAS analyses will be performed to assess associations between genotype and endpoint, either within each clinical study or within groups of clinical studies for the same indication that were genotyped using the same platform (e.g. VEG102616, VEG105192 and VEG107769). Association analyses will be conducted using ProbABEL[Aulchenko, 2010], mach2qtl / dat[Li, 2010], PLINK [Purcell, 2007] or equivalent software. In the next step, the effect size estimates and standard errors (SEs) for genetic variants thus obtained from each GWAS analysis will be meta-analyzed across studies using fixed effect inverse variance weighting, as implemented in METAL (http://genome.sph.umich.edu/wiki/METAL_Program) or equivalent software. After meta-analysis, top-hit SNP markers (defined by P-value threshold or rank) will be annotated.

Additional association may be tested using race stratified analysis, in order to explore whether there are race-specific GWAS signals, or whether the effect size of any GWAS top-hit from the overall analysis differs between sub-populations such as whites versus East Asians. In addition, GWAS findings from the primary analysis in pazopanib-treated subjects will be tested for baseline ALT and in subjects in the control arm (placebo or active) if sufficient data are available. This would help to evaluate whether the genetic association with ALT elevation is pazopanib specific or not. Additionally as specified in Section 4.2, the analyses may be expanded to make use of recently published analyses methods such as using the gene-enriched and pathway-based analyses to study how single gene or genes within the same biological cascade act as one function unit to affect ALT. Other exploratory analyses may use multivariate analysis and a longitudinal framework to study how the genetic variation could affect the trajectory pattern (change) of ALT. Multivariate analyses (PCA, factor and clustering analysis etc.) may be pursued to create composite phenotypes, which can take advantage of combining available information from measurements across multiple time points. The longitudinal data analysis methods are capable of increasing flexibility to model how genetic variants may associate with ALT level changes as a function of time. In theory, these novel analytic methods may increase the power to detect genetic association signals. Finally, expression quantitative trait loci (eQTLs) of mRNA and miRNA, and methylation QTLs may be applied to leverage the knowledge on the putative function of identified genetic variants from GWAS analysis [Gamazon, 2010]. Additional exploratory analyses may be conducted at the discretion of the team.

6. DATA HANDLING CONVENTIONS

6.1. Premature Withdrawal and Missing Data

If a subject contributed data to the analysis population and also provided PGx consent and a DNA sample, data from the subject will be included in the PGx analyses outlined here. If a subject withdrew from the clinical study but did not withdraw consent for PGx research, the subject's information may still be used in the PGx analyses.

Primary **candidate gene analyses** will not use imputed values for dependent variables or covariates, but for the **GWAS analysis**, missing genotype data will be imputed (Section 5.4.2). This is because the model for genotype imputation is based on a realistic population genetic model of the underlying data generating process, and has been empirically validated in many independent studies.

7. REPORTING CONVENTIONS

7.1. Demographic, Baseline and Endpoint Characteristics

The number of subjects included in each analysis population will be reported.

Demographics of each analysis population, if necessary further split by phenotypic covariates will be summarized. In general, continuous variables will be summarised by the mean or median, standard deviation or interquartile range, and minimum and maximum. Categorical variables will be summarised with frequency counts and percentages.

In **GWAS analysis**, summary demographics will be used to compare the PGx analysis populations against the respective ITT populations.

7.2. Summary of Genetic Association Analysis Results

For the candidate gene analyses, the distribution of observed P-values and the null distribution estimated from permutations will be reported, and the two distributions using histograms or kernel density estimates will be compared. For each significant association, the candidate gene and alleles, cross-tabulated with case/control status, along with association P-value, effect size (odds ratio) and standard error from the analysis adjusted for race and ethnicity will be reported.

For the GWAS analysis, Manhattan plots and Quantile-Quantile (QQ) plots will be used to visualize P-values at the whole genome scale within each clinical study and the meta-analysis. For each significant or suggestive (defined by cut-off P-value or top ranks) association, the genetic variant and alleles, allele frequencies, effect size and SE within each clinical study and meta-analysis effect size, SE and P-value will be reported. Forest plots will be used to visualise the study-level results. Results by clinical study will be annotated by whether the genetic variant was typed or imputed, and a metric for quality of imputation. Within-study association will be graphed using e.g. box-and-strip-charts to plot a continuous endpoint against genotype for all subjects in each study, or Kaplan-Meier estimates of survival fraction by genotype for all subjects in each study. Biological

functions and pathway information will be checked for genes harboring or near top-hit SNPs to see if any of these nominated genes is directly or indirectly related to liver toxicity.

8. REFERENCES

Aulchenko YS, Stuchlin MV, and van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics.* 2010;11:134.

Dudbridge F, Gusnanto A. Estimation of Significant Thresholds for Genomewide Association Scans. *Genetic Epidemiology.* 2008;32:227-34.

Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ. SCAN: SNP and copy number annotation. *Bioinformatics.* 2010;26:259-62.

Goodman V, Wang K., Pandite L, Watkins P.B. Incidence and management of hepatic toxicity in pazopanib-treated patients. *Annals of Oncology.* 2010;21:viii282.

Harper AR, Topol EJ. Pharmacogenomics in clinical practice and drug development. *Nature Biotechnology.* 2012;30:1117-24.

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.

Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, Mangino M, Albrecht E, Wallace C, Farrall M, Johansson A, Nyholt DR, Aulchenko Y, Beckmann JS, Bergmann S, Bochud M, Brown M, Campbell H; EUROSPAN Consortium, Connell J, Dominiczak A, Homuth G, Lamina C, McCarthy MI; ENGAGE Consortium, Meitinger T, Mooser V, Munroe P, Nauck M, Peden J, Prokisch H, Salo P, Salomaa V, Samani NJ, Schlessinger D, Uda M, Völker U, Waeber G, Waterworth D, Wang-Sattler R, Wright AF, Adamski J, Whitfield JB, Gyllensten U, Wilson JF, Rudan I, Pramstaller P, Watkins H; PROCARDIS Consortium, Doering A, Wichmann HE; KORA Study, Spector TD, Peltonen L, Völzke H, Nagaraja R, Vollenweider P, Caulfield M; WTCCC, Illig T, Gieger C. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLos Genetics.* 2009;5:e1000504.

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.

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology.* 2010;34:816-34.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics.* 2008;9:356-69.

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

Pandite L, Goodman V., Botbyl J., Suttle A.B., Gauvin J., Dar M. Trough PK concentration, age of patient, and baseline alt are important factors in predicting alt elevations in pazopanib treated mrcc patients. *Annals of Oncology.* 2010;21S8:viii286.

Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLoS Genetics.* 2006;2:e190.

Price AL, Patterson NJ, Plenge R.M, Weinblatt ME, Shadick N.A, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics.* 2006;38:904-9.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics.* 2007;81:559-75.

Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, de Bakker PI. ve amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genetics.* 2012;44:291-6.

Spraggs CF, Parham LR, Hunt CM, Dollery CT. Lapatinib-induced liver injury characterized by class II HLA and Gilbert's syndrome genotypes. *Clinical Pharmacology & Therapeutics.* 2012;91:647-52.

Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, Barrios CH, Salman P, Gladkov OA, Kavina A, Zarbá JJ, Chen M, McCann L, Pandite L, Roychowdhury DF, Hawkins RE. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *Journal of Clinical Oncology.* 2010;28:1061-8.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC et.al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010;466:707-13.

van der Graaf WT, Blay JY, Chawla SP, Kim DW, Bui-Nguyen B, Casali PG, Schöffski P, Aglietta M, Staddon AP, Beppu Y, Le Cesne A, Gelderblom H, Judson IR, Araki N, Ouali M, Marreaud S, Hodge R, Dewji MR, Coens C, Demetri GD, Fletcher CD, Dei Tos AP, Hohenberger P; EORTC Soft Tissue and Bone Sarcoma Group; PALETTE study group. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet.* 2012;379:1879-86.

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.

Xu CF, Reck BH, Xue Z, Huang L, Baker KL, Chen M, Chen EP, Ellens HE, Mooser VE, Cardon LR, Spraggs CF, Pandite L. Pazopanib-induced hyperbilirubinemia is associated with Gilbert's syndrome UGT1A1 polymorphism. *British Journal of Cancer.* 2010;102:1371-7.

Yang J, Ferreira T, Morris AP, Medland SE, GIANT Consortium, DIAGRAM Consortium, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME and Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics.* 2012;44:369-75.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics.* 2012;44:821-4.

9. APPENDIX

Table 1 Number of concurrent ALT and bilirubin elevation cases in the 16 clinical studies for the candidate gene analysis

Study ID	Phase / Indication	Study Description	Cases, N
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	1
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	1
VEG107200	I / Hepatocellular Cancer	A Phase I, Open-Label, Dose Escalation, Multi-Center Study of pazopanib (GW786034) in Adults Subjects with Hepatocellular Cancer	1
VEG109607	I / Solid Tumours	A Phase I Study of Pazopanib in Combination with Either Erlotinib or Pemetrexed in Patients with Advanced Solid Tumors	2
VEG102616	II / RCC	A Phase II Study of GW786034 Using a randomized Discontinuation Design in Subjects with Locally Recurrent or Metastatic Clear-Cell Renal Cell Carcinoma	1
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	1
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	1
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	3
VEG113078	II / RCC	GW786034 A Phase II study to evaluate efficacy and safety of pazopanib and sunitinib in Advanced RCC patients from Asian (Substudy to VEG108844)	3
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	1
VEG105192	III / RCC	A Randomized, Double-blind, Placebo-controlled, Multi-center Phase III Study to Evaluate the Efficacy and Safety of Pazopanib (GW786034) Compared to Placebo in Patients with Locally Advanced and/or Metastatic Renal Cell Carcinoma	4
VEG107769	III / RCC	An open-label extension study to assess the safety and efficacy of pazopanib in subjects with renal cell carcinoma previously enrolled on protocol VEG105192	2
VEG108844	III / RCC	A study of Pazopanib versus Sunitinib in the Treatment of Subjects with Locally Advanced and/or Metastatic Renal Cell Carcinoma	2
VEG110655	III / Ovarian Cancer	A Phase III Study to Evaluate the Efficacy and Safety of Pazopanib Monotherapy Versus Placebo in Women Who Have not Progressed after First Line Chemotherapy for Epithelial Ovarian, Fallopian Tube, or Primary Peritoneal Cancer	3
VEG110727	III / STS	A randomized double blind phase III trial of Pazopanib versus placebo in patients with soft tissue sarcoma whose disease has progressed during or following prior therapy	3
VEG113387	III / RCC	A randomized, double-blind, placebo-controlled phase III study to evaluate the efficacy and safety of pazopanib as adjuvant therapy for subjects with localized or locally advanced RCC following nephrectomy	3
Total	16 studies		32

Table 2 List of clinical studies for the GWAS analysis*

Study	Phase / Indication	ITT		PGx population[1]		Race / Ethnicity Ratio White: Asian: Other[2]	Illumina array chip used for GWAS
		Total	Pazopanib treated	Total	Pazopanib treated		
VEG102616	II / RCC	225	225	164	164	116: 25: 23	Human1M
VEG105192	III / RCC	435	290	278	182	131: 16:35	Human1M
VEG107769	III / RCC	80	79**	57	57	33: 8:16	Human1M
VEG108844	III / RCC	927	464	590	302	213: 61: 28	HumanOmni5-quad
VEG113078	II / RCC	183	93	146	77	0: 76: 1	HumanOmni5-quad
VEG110727	III / STS	369	246	125	92	57: 29: 6	HumanOmni5-quad
VEG110655	III / Ovarian cancer	940	477	676	346	246: 79: 21	Omni-Exp+Exome
VEG114012	II/ Ovarian cancer	145	73	109	53	0: 53: 0	Omni-Exp+Exome
Total	**8 studies**	3304	1947	2145	1273	**796: 347: 130**	**three Illumina platforms**

[1] PGx population is defined as subjects who were consented for the PGx research and had sufficient DNA sample for genotyping.
[2] Counts are based on PGx pazopanib treated population. White = non-Hispanic Whites; Asian = East and South-East Asian; Other = any other subjects combined
*The counting information was obtained from PGx308, PGx420, PGx421, PGx6202 respectively.
** One subject was duplicated in VEG107769 and VEG105192 pazopanib treated arm, and removed from VEG107769.

Table 3 Summary of genetic markers in candidate gene analysis for concurrent ALT and bilirubin elevation

Gene	Markers	Rationale
HLA-A *HLA-B* *HLA-C* *HLA-DRB1* *HLA-DQA1* *HLA-DQB1* *HLA-DPB1*	4-digit HLA allele	Associated with drug induced hepatotoxicity
UGT1A1	*28 (rs8175347) *6 (rs4148323)	Associated with bilirubin elevation in pazopanib treated subjects with RCC
CYP3A4	*1B (rs2740574)	Pazopanib metabolism / disposition
CYP3A5	*3 (rs776746)	
CYP2C8	*3 (rs10509681)	
ABCB1	1236 C>T (rs1128503) 2677 G>T/A (rs2032582) 3435C>T (rs1045642)	
ABCG2	421C>A (Q141K) (rs2231142) 34 G>A (V12M) (rs2231137) 869C>T (Q126X) (rs72552713)	

Table 4 Genetic variants for candidate gene lookups in the GWAS analysis

Gene	SNP reference number	Polymorphism	Rationale
HLA-A *HLA-B* *HLA-C* *HLA-DRB1* *HLA-DQA1* *HLA-DQB1* *HLA-DPA1* *HLA-DPB1*		HLA alleles by imputation	Associated with drug induced hepatotoxicity
UGT1A1	rs8175347 rs4148323	*28 *6	Associated with bilirubin elevation in pazopanib treated subjects with RCC
CYP3A4	rs2740574	*1B	Pazopanib metabolism / disposition
CYP3A5	rs776746	*3	
CYP2C8	rs10509681	*3	
ABCB1	rs1128503 rs2032582 rs1045642	1236 C>T 2677 G>T/A 3435C>T	
ABCG2	rs2231137 rs2231142 rs72552713	34G>A (V12M) 421C>A (Q141K) 869C>T (Q126X)	
NR1I2	rs3814055	-25385C>T	CYP3A4 expression regulation
NR1I3	rs2307424 rs2307418 rs4073054	5719C>T 7738A>C 7837T>G	
VEGFR2 (KDR)	rs2071559 rs1870377 rs2305948 rs34231037	-604T>C Q472H (1718T>A) V297I (889G>A) C482R	Pazopanib target
VEGFR3 (FLT4)	rs307826 rs307821	T494A R1324L	
VEGFA	rs699947 rs833061 rs1570360 rs2010963 rs3025039	-2578A>C -1498C>T -1154G>A -634G>C 936C>T	Ligand of pazopanib target
HFE	rs2858996 rs707889 rs1800562	C282Y	May be associated with ALT elevation in pazopanib treated subjects with RCC