Division: Worldwide Development
Retention Category: GRS019
Information Type: Pharmacogenetics Reporting and Analysis Plan (Genetic)

Title: Reporting and Analysis Plan for VEG116054 (PGx447): An Investigation of Associations between Genetic Markers and Pazopanib (GW786034)-related Diarrhoea in Patients with Renal Cell Carcinoma from Studies VEG102616, VEG105192, and VEG107769

Compound Number: GW786034

Effective Date: 24-OCT-2011

Description: This document defines the intended analysis strategy for the reporting of pharmacogenetic data from three Pazopanib protocols VEG102616, VEG105192, and VEG107769.

Diarrhoea is one of the most common adverse events in patients treated with pazopanib, and significantly affects patients' quality of life and treatment outcomes. The mechanism of underlying diarrhoea in pazopanib treated patients is unclear and no predictive factor has been identified. The objective of this analysis is to investigate if genetic markers may explain the risk of diarrhoea observed in patients treated with pazopanib. To accomplish this, genetic markers from 38 candidate genes, which are involved in the pharmacokinetic and pharmacodynamic pathway of pazopanib, or may be related to drug induced diarrhoea, and markers from Illumina Human1M DNA Analysis BeadChip will be evaluated.

Subject: Pharmacogenetics (PGx), pazopanib (GW786034), SNP, Case-Control analysis, genome-wide association study (GWAS), candidate gene analysis, renal cell carcinoma (RCC), and diarrhoea.

Author's Name, Title and Functional Area:

[redacted] Investigator Statistical Genetics, Quantitative Sciences

[redacted] Investigator Statistical Genetics, Quantitative Sciences

[redacted] Investigator Therapy Area Applied Genetics, Genetics

Approved through email by:

[redacted]	24th October 2011
SM Director, Therapy Area Applied Genetics, Genetics	Date
[redacted]	21st October 2011
Senior Scientific Investigator, Statistical Genetics, Quantitative Sciences	Date

Distribution List:

[redacted]	Dir Statistics & Programming, Oncology TA
[redacted]	Manager, Statistics, Oncology TA
[redacted]	Dir Clinical Development, MDC Clinical Oncology
[redacted]	Manager, GCSP Oncology Safety Evaluation & Risk Management Group
[redacted]	Investigator, Statistical Genetics, Quantitative Sciences
[redacted]	Senior Scientific Investigator, Statistical Genetics, Quantitative Sciences

TABLE OF CONTENTS

PAGE

ABBREVIATIONS & GLOSSARY....5

1. INTRODUCTION....7

2. STUDY OBJECTIVES AND ENDPOINT....8
   - 2.1. Study Objective....8
   - 2.2. Study Endpoints....8

3. STUDY DESIGN....9

4. PLANNED ANALYSES....10

5. SAMPLE SIZE CONSIDERATIONS....10
   - 5.1. Retrospective Power Calculations....11

6. ANALYSIS POPULATIONS....12

7. GENERAL CONSIDERATIONS FOR DATA ANALYSES....12
   - 7.1. Dependent Variable (Endpoint Variable)....12
   - 7.2. Independent Variables....13
   - 7.3. Genetic Markers....13
     - 7.3.1. Candidate Genes and Genetic Markers....13
     - 7.3.2. Genetic Markers from GWAS....14
   - 7.4. Marker Map....14
   - 7.5. Examination of Subgroups....14
   - 7.6. Multiple Comparisons and Multiplicity....14

8. DATA HANDLING CONVENTIONS....15
   - 8.1. Premature Withdrawal and Missing Data....15

9. STUDY POPULATION....15
   - 9.1. Disposition of Subjects....15
   - 9.2. Demographic and Baseline Characteristics....15

10. GENETIC ANALYSIS....15
    - 10.1. Data Quality Control....16
      - 10.1.1. Genotype Quality Control....16
      - 10.1.2. Subject Quality Control....16
    - 10.2. Hardy-Weinberg Equilibrium (HWE) Analysis....16
    - 10.3. Linkage Disequilibrium Analysis....17
    - 10.4. Ancestry Estimates....17
    - 10.5. Evaluation of Genotypic Association....17
      - 10.5.1. Summary of Association Analysis Results....18

11. INTERPRETATION AND REPORTING OF ASSOCIATION ANALYSES RESULTS....19
    - 11.1. Focus on Strongest Exploratory Results....19

12. REFERENCES....20

13. APPENDIX 21
13.1. Diarrhoea PGx Candidate Gene List 21

ABBREVIATIONS & GLOSSARY

AE	Adverse event
ADME	Absorption, distribution, metabolism, and excretion
Allele	Any 1 of 2 or more alternative forms of a gene or DNA sequence at the same locus. One allele is inherited from each parent
ANCOVA	Analysis of Covariance
Candidate gene	A gene hypothesized to influence a specified event (e.g., response to drug therapy)
CC	Case-control
DNA	Deoxyribonucleic acid
FET	Fishers Exact Test
Genetic Locus	A region on a chromosome, can be an entire gene, part of a gene or a segment of DNA of unknown function
Genetic Marker	Identifiable physical location on a chromosome at which inheritance can be monitored
Genotype	A pair of alleles, one having come from each biological parent that characterizes an individual at a specified genetic locus. Different combinations can produce different traits/characteristic
HWE	Hardy Weinberg Equilibrium (HWE) analysis provides a measure of the association between two alleles at an individual locus. Markers showing evidence of departures from HWE can be investigated thoroughly for laboratory errors or other causes of departure from equilibrium.
IDSL	Integrated Data Standards Library
ITT	Intent to treat.
LD	Linkage Disequilibrium (LD) is a measure of association between alleles at different loci.
OR	Odds ratio, the ratio of the odds of cases having the suspect genotype to the odds of controls having the suspect genotype (or: the odds of being a case given carriage of the suspect genotype divided by the odds of being a case given the absence of the suspect genotype).
PGx population	Consists of subjects in the clinical intent to treat (ITT) population who consent to a genetic study and provide an adequate DNA sample for genotyping.

PGx analysis population	The collection of subjects from the clinical studies under investigation who provided written informed consent for genetic research, provided a blood sample for genotyping that was successfully genotyped for at least one of the genetic markers under study, and passed a subject data quality control check.
Polymorphism	A genetic locus at which there is a difference in DNA sequence among individuals (i.e. which has 2 or more alleles in the population)
QC	Quality Control
RAP	Reporting and analysis plan
Risk Genotype	The genotype of an genetic marker that is associated with diarrhea case status
SNP	Single nucleotide polymorphism

Trademark Information

Trademarks of the GlaxoSmithKline group of companies
VOTRIENT™

Trademarks not owned by the GlaxoSmithKline group of companies
None

1. INTRODUCTION

Pazopanib (VOTRIENT™, GlaxoSmithKline) is an oral angiogenesis inhibitor targeting vascular endothelial growth factor receptors (VEGFR) -1, -2, and -3, platelet-derived growth factor receptors, and the stem cell factor receptor, c-Kit. Pazopanib is approved in multiple countries for the treatment of advanced renal cell carcinoma (RCC). Pazopanib has demonstrated clinical activity against multiple tumour types, and its safety profile is generally acceptable and tolerable. The most commonly observed adverse reactions (≥20%) in the 586 patients with RCC were diarrhoea, hypertension, hair color change, nausea, fatigue, anorexia, and vomiting. The prevalence of diarrhoea in the phase III RCC trial (N=290) was 52% for all grades and ~4% for grade 3 and above in pazopanib treated subjects, compared to 9% for all grades and <1% for grade 3 and above in placebo treated subjects (N=145), respectively [Sternberg, 2010] .Up to date, the mechanism of pazopanib related diarrhoea is unclear. No biomarker has yet been identified as a predictor for pazopanib-related diarrhoea.

Diarrhoea is one of the most common side effects of cancer treatment and can be a serious and debilitating complication. Careful analysis and understand of the causative factors can help on a more accurate management and early intervention, and possibly prevent severe complications for patients.

Several pathophysiological mechanisms for potential drug-induced diarrhoea have been hypothesized. Evidence suggests that drug induced diarrhoea is probably a multifactorial process that results in absorptive and secretory imbalances in the bowel [Abraham, 2007]. Cancer treatment (ie. chemotherapy drugs) can cause diarrhoea by damaging immature epithelial cells in the crypt which results in subsequent mature enterocytes being functionally impaired, decreasing their absorptive capacity or by disrupting the integrity of the intestinal mucosa through inflammatory and ulcerated lesions. Additionally, a disruption to normal active ion absorption in the small intestine can lead to an excess of water and electrolytes [Stein, 2010]. Identification of genes that act on these pathophysiological pathways may help our understanding of the pazopanib-related diarrhoea mechanism and, thus, patient susceptibility to it.

The purpose of this investigation is to determine if observed diarrhoea events seen in pazopanib treated RCC patients may be associated with genetic risk factors. Genetic markers in 38 candidate genes and markers from Illumina Human1M DNA Analysis BeadChip will be evaluated. See Section 13.1 for specific information on the candidate genes investigated.

Patients who consented and provided a DNA sample from the following clinical trials are included in this genetic investigation: VEG102616, VEG105192, and VEG107769.

This document describes the analyses planned to explore the impact of genetic variation on diarrhoea in pazopanib treated subjects through a primary analysis on defined diarrhoea case-control (CC) endpoint, and a secondary analysis on ordinal endpoint by using diarrhoea toxicity grade levels. As these data are pooled from three clinical trials, special consideration will be taken with respect to relevant clinical factors. All analyses are exploratory with respect to the genetic markers being evaluated.

This document contains a brief summary of the study design followed by descriptions of the analysis populations, analysis variables, statistical methods and possible ways to summarise the data that will comprise the analysis.

2. STUDY OBJECTIVES AND ENDPOINT

2.1. Study Objective

The objective of this analysis is to investigate if genetic markers from 38 candidate genes or other markers from across the genome are associated with incidence of diarrhoea in RCC patients treated with pazopanib in studies VEG102616, VEG105192 and VEG107769.

2.2. Study Endpoints

The primary endpoint is the diarrhoea outcome status defined as:

- **Case** - a patient who experienced a minimum of Grade 2 diarrhoea based on the National Cancer Institute Common Toxicity Criteria v3.0 (Table 2-1) during pazopanib treatment.

- **Control** - a patient who received pazopanib for a minimum of 42 days, but never developed diarrhea. Forty-two days was the median time to first diarrhea event in patients treated with pazopanib in the studies.

  Note that patients who experienced diarrhoea episodes at grade 1 are not included in the case-control analysis.

The secondary endpoint is diarrhoea toxicity grade levels from 1 to 5 (Table 2-1) and grade 0 for the control described above, defined as an ordinal categorical variable. As the number of subjects who experienced high diarrhoea toxicity grade (Grade 3+) is small, it may be possible to pool subjects with Grade 2-5 into a single category for the purposes of this analysis.

Table 2-1 NCI's Common Toxicity Criteria v3.0 for Diarrhoea*

Grade	Description
1	Increase of <4 stools per day over baseline; mild increase in ostomy output compared to baseline
2	Increase of 4 – 6 stools per day over baseline; IV fluids indicated <24hrs; moderate increase in ostomy output compared to baseline; not interfering with activities of daily living
3	Increase of ≥7 stools per day over baseline; incontinence; IV fluids ≥24 hrs; hospitalization; severe increase in ostomy output compared to baseline; interfering with activities of daily living
4	Life-threatening consequences (e.g., hemodynamic collapse)
5	Death

1. *Diarrhoea includes diarrhoea of small bowel or colonic origin, and/or ostomy diarrhoea.

3. STUDY DESIGN

Data included in this analysis were derived from clinical studies VEG102616, VEG105192, and VEG107769, all pazopanib monotherapy studies in subjects with RCC.

The clinical data used will be from the Integrated Safety Summary (ISS) for these three studies. The ISS data came from the SAC analyses for VEG102616 and VEG105192, with data cutoffs of April 2008 and May 2008 respectively. Interim data for VEG107769 was included in ISS with a date cutoff of May 2008.

A brief description of the studies, with the number of subjects available for genotyping is provided in Table 3-1.

Table 3-1 Description of the Clinical Studies

Study	ITT Population1	Pazopanib treated ITT Population1	PGx Population2	Study Design	Description
VEG102616	225	225	164	A phase II, pazopanib monotherapy, randomized discontinuation design study	To determine the efficacy and safety of pazopanib in subjects with locally recurrent or metastatic clear-cell renal cell carcinoma. Subjects received pazopanib administered orally at a daily dose of 800 mg.
VEG105192	435	290	182	A phase III, randomized, double blind, placebo-controlled, multi-center study	Eligible patients were stratified and randomized in a 2:1 ratio to receive either 800 mg of pazopanib once daily or matching placebo.
VEG107769	71	70	51	A phase III, open-label, pazopanib monotherapy, multi-center study	An open-label, extension study, to evaluate safety and efficacy of pazopanib in patients who randomized into study VEG105192 and experienced progressive disease after receiving placebo as the study treatment. Subjects receive 800 mg pazopanib once daily.
Total	731	585	397		

1. The ITT Population consists of subjects who were enrolled in studies with intent to treat.
2. The PGx Population is comprised of subjects from the ITT population who provided written informed consent for PGx research and an adequate DNA sample for genotyping.

4. PLANNED ANALYSES

Primary analysis is to test association of case-control endpoint with each genetic marker. Secondary analysis is to test association of ordinal endpoint with each genetic marker. The analysis will be conducted using pazopanib treated patients of all racial groups pooled, and will use genetic ancestry information (derived from Principal Component analysis of the genotype data) to control for potential population stratification.

5. SAMPLE SIZE CONSIDERATIONS

The clinical studies being used in this analysis were not prospectively designed to address genetic research hypotheses and, thus, did not benefit from prospective sample size calculations. Instead of calculating the sample size necessary to attain a certain statistical power, the anticipated sample size will be used to estimate power under a variety of potential experimental scenarios and genetic models.

5.1. Retrospective Power Calculations

The power to detect a significant statistical genetic association is dependent upon several factors, such as the population prevalence of the event, the frequency of the risk genotype in cases, the size of the genetic effect, and the number of subjects analyzed. Basic retrospective power calculations were conducted to inform what kinds of genetic effects could be found given the observed numbers of diarrhoea cases in the clinical studies. These power calculations might be over-estimated given the assumption that no covariates are included in the model and that the causal variant itself will be assessed (as opposed to a variant in partial linkage disequilibrium with the causal variant).

Table 5-1 presents the subject count at each toxicity grade of diarrhoea in pazopanib treated ITT and PGx population. Power was calculated for diarrhoea case-control status, based on the expected number of RCC patients in pazopanib treated PGx population. The power was estimated assuming a 2 degree of freedom genotypic test conditional on the underlying genetic model being additive, dominant or recessive, respectively. These calculations assumed a type I error rate of 3.1×10^{-4}, the significance threshold with Bonferroni correction for Tier 1 markers (see Section 7.6 for the definition of Tier 1 markers). The plots show the relationship between statistical power and the odds ratio (OR) of the two homozygotes (Figure 5-1). The different curves correspond to power estimates assuming six different risk allele frequencies (RAFs).

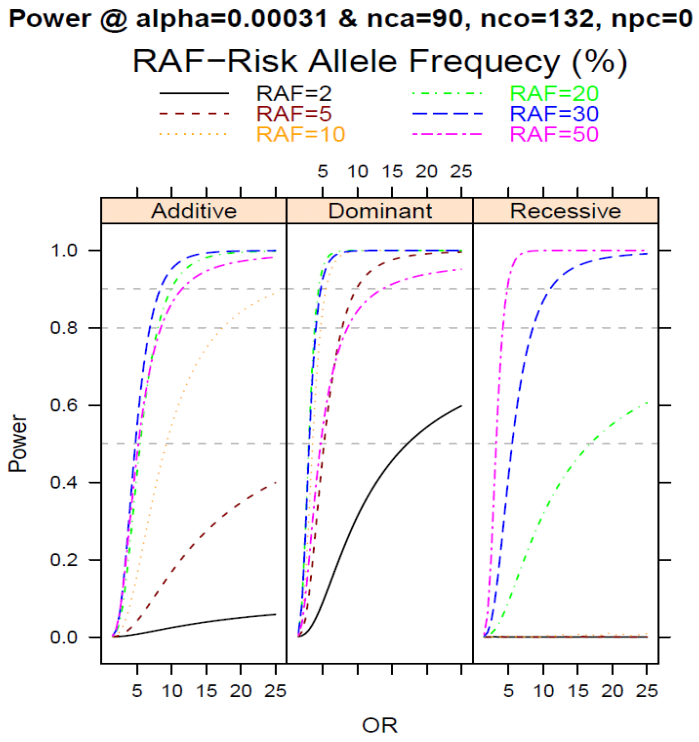
Table 5-1 Summary of diarrhoea for pazopanib treated subjects in VEG102616, VEG105192 and VEG107769

Diarrhoea AE grade	ITT (%)	PGx population (%)
0*	266 (45%)	156 (39%)
1	202 (34%)	151 (38%)
2	95 (16%)	78 (20%)
≥3	21 (3.6%)	12 (3%)
Missing data	2 (0.3%)	0
Total	586	397

1. * Without consideration of treatment time for pazopanib

With 90 diarrhoea cases and 132 controls as defined in Section 2.2, there are many combinations of genotypic models, risk allele frequencies and odds ratios, that may lead to good statistical power (>80%) to detect an effect, especially if the dominant or additive models are true and if the risk allele is relatively common in the study population (i.e., MAF≥20% with a genetic effect OR ≥ 4). The power to detect significant associations with SNPs from GWAS will be considerably reduced in comparison given approximately 1 million markers tests.

Figure 5-1 Power Estimation



6. ANALYSIS POPULATIONS

The Genetic Analysis Population will consist of subjects enrolled in clinical studies VEG102616, VEG105192 and VEG107769 who provided written informed consent for genetics research, provided a blood sample for genotyping and were successfully genotyped for at least one of the genetic markers under study, have valid phenotype data and pass subject QC as described in Section 10.1.2.

7. GENERAL CONSIDERATIONS FOR DATA ANALYSES

It is anticipated that after an initial review of the results of the analyses described here, there may be need for additional follow-up analyses to be conducted. These will be discussed, defined, and agreed by the authors of this RAP and other relevant parties at that time. In that case, there will be no formal amendments to this RAP; any additional follow-up analyses will be described in the Genetic Study report/synopsis.

7.1. Dependent Variable (Endpoint Variable)

The dependent variables will:

- Diarrhoea case-control status, and (as defined in Section 2.2).
- Diarrhoea toxicity levels (defined an ordinal categorical variable).

7.2. Independent Variables

The following variables will be assessed for inclusion as covariates whilst testing for genotype effects.

- Study
- Gender (M/F)
- Age (yrs)
- Baseline body mass index (BMI)
- Baseline patients performance status, ECOG score
- Baseline neutrophil count
- Ethnicity/Ancestry information derived as outlined in Section 10.4

7.3. Genetic Markers

7.3.1. Candidate Genes and Genetic Markers

Thirty-eight candidate genes were selected for inclusion in this investigation (Table 7-1). These include genes that are involved in the pharmacokinetic and pharmacodynamic pathway of pazopanib, or may be related to the presumed mechanism of diarrhoea. 1380 SNPs including 53 functional genetic markers, which map to these genes, will be evaluated. For detailed information of the candidate genes and markers, see Appendix (section 13.1).

Table 7-1 Diarrhoea Candidate Genes List

Group	Gene Name	Number of genes
Pazopanib ADME	CYP3A4/5 CYP2C8 NR1I2 ABCB1 ABCG2	6
Pazopanib mode of action	VEGFR1 / 2 / 3 PDGFR α / β c-kit VEGFA	7
Diarrhoea	SLC6A4 (5-HT) CFTR CLCA1 SLC12A2 ATP1A1 ADCY3 / ADCY6 NOS3 SLC9A3 SLC5A1 SLC26A3 SCNN1A / SCNN1B / SCNN1G ANO1 TNF IFNG IL1B / 4 / 6 / 8 / 10 / 12A / 12B / 13	25
Total		38

7.3.2. Genetic Markers from GWAS

In addition to the candidate gene markers, analysis will be performed using the rest of genetic markers from Illumina Human1M DNA Analysis Beadchip.

7.4. Marker Map

The NCBI 36 map will be used in this analysis

7.5. Examination of Subgroups

No subgroup has been identified at this time for independent analysis.

7.6. Multiple Comparisons and Multiplicity

A set of p-value thresholds will be used to claim significant association. To establish these thresholds, control of the family-wise type I error for multiple tests will be calculated as described below. All genetic markers are categorized into three tiers based on the prior biological information. The first tier includes 53 functional markers from the candidate genes. The second tier consists of the rest of the markers (~1300 SNPs) in the candidate genes. The third tier will comprise all other Illumina 1M markers that do not map to the 38 candidate genes (~1M markers). A family-wise error rate of 0.05 is allocated evenly to the three tiers, that is, 0.0167 (determined from 0.05/3). The

significant p-value thresholds for the first, the second and third tiers are calculated using the Bonferroni correction, which is 0.0167 / (number of markers), for each analysis at each tier. Linkage disequilibrium among markers within each category will be taken into account and an effective number of tests in Tier 1 & 2 can be estimated for the purpose of the multiple test adjustment [Moskvina, 2008]. Once the actual number of markers to be analyzed within each category has been determined, a per-marker significance level will be calculated and applied to the statistical test results.

Genetic markers that satisfy nominal p-value thresholds of ≤ 0.05, ≤ 0.01, and $\leq 10^{-4}$ in the first, second and third tiers, respectively, will also be summarized as suggestive evidence for association; which will be further evaluated in future studies.

8. DATA HANDLING CONVENTIONS

8.1. Premature Withdrawal and Missing Data

If a subject is included in the Genetics Analysis Population, as defined in Section 6 then the subject will be included in the analyses outlined here. If a subject withdrew from the study but did not withdraw consent for genetic research, the subject's information will still be used in these analyses.

Missing data will not be imputed in the analysis.

9. STUDY POPULATION

9.1. Disposition of Subjects

The number of subjects in the PGx Analysis Population will be summarized overall and by study.

9.2. Demographic and Baseline Characteristics

Summaries of baseline demographics will be produced as needed. In general, categorical data will be summarized using frequency counts and percents, and continuous data will be summarized using means, standard deviations, minimums, medians, and maximums, overall and by grade of diarrhoea experienced. These summary statistics will be inspected visually for any concerning imbalances.

10. GENETIC ANALYSIS

The Case-Control analyses will be performed on the binary outcome measure (diarrhoea status), using logistic regression analysis; the ordinal data analyses will be performed on ordinal categorical diarrhoea toxicity grade levels, using proportional odds model, if the marker meets the p-value threshold for suggestive evidence in the primary anlaysis. The analysis strategies that will be employed are outlined in this section.

10.1. Data Quality Control

Prior to the genetic association analysis, quality control (QC) will be conducted on the genetic markers themselves and on the subjects utilizing the genetic marker data.

10.1.1. Genotype Quality Control

The objective of genotype QC analysis is to define the genetic markers that are of high quality and suitable for use in genetic association analyses. For this investigation, genotype data will be extracted from the existing data in the previous pazopanib PGx studies [see GlaxoSmithKline Document Number HJ2009/00002/00 and GlaxoSmithKline Document Number RM2008/00332/00].

10.1.2. Subject Quality Control

The subjects for this investigation have been QC'ed previously [see GlaxoSmithKline Document Number HJ2009/00002/00 and GlaxoSmithKline Document Number RM2008/00332/00] .

10.2. Hardy-Weinberg Equilibrium (HWE) Analysis

Departure from Hardy-Weinberg equilibrium (HWE) is a measure of the association between two alleles at an individual locus. A diallelic marker locus is in HWE if the frequencies of the genotypes 11, 12 and 22 are p^2, $2pq$ and q^2 respectively, where p and $q = 1 - p$ are the frequencies of the alleles 1 and 2 respectively. The expected frequencies of the three genotypes will be obtained from the observed allele frequencies, on the assumption that the locus is in HWE. Significant departure from HWE may indicate a laboratory genotyping error and population genetic imbalance.

HWE analysis will be performed on all markers and will be conducted using the White subjects sub-population. Departure from HWE will be tested using an exact test, by considering the distribution of genotypes conditional on observed allele frequencies. No markers will be removed prior to association analysis based on their HWE P-values. Markers showing substantial evidence of departure from HWE will be investigated thoroughly for potential laboratory errors or other causes of departure from equilibrium and, if the cause remains undetermined, they may be removed from the list of associated markers. Results pertaining to such markers will be interpreted cautiously.

10.3. Linkage Disequilibrium Analysis

Linkage Disequilibrium (LD) analysis will be conducted to measure the association between alleles at different loci. The LD between two markers can be given by $D_{AB} = p_{AB} - p_A p_B$, where p_A is the allele frequency of A allele of the first marker, p_B is the allele frequency of B allele of the second marker, and p_{AB} is the joint frequency of alleles A and B on the same haplotype. Linkage disequilibrium will be tested using a chi-square test. To limit the number of comparisons, LD analysis will be conducted only on markers within the same gene, or gene region.

Linkage Disequilibrium analysis will be conducted in the White subjects sub-population. LD in other subgroups of subjects may be investigated, as appropriate.

Pairwise LD in the form of r^2 (the squared correlation of genotypes of each pairwise SNP combination) will be calculated for SNPs of interest.

10.4. Ancestry Estimates

Genetic ancestry estimates will be obtained by principal component analysis using EIGENSOFT version 2.0 BETA software package (http://genepath.med.harvard.edu/~reich/Software.htm). Genetic data of all PGx subjects who passed QC from clinical studies, VEG102616, VEG105192 and VEG107769, will be seeded with genetic data from HapMap subjects which includes 60 Yorubans, 60 Caucasians and 90 Asians (45 from China and 45 from Japan). Diallelic autosomal SNPs from the GWAS in common with SNPs from HapMap Release 23 will be used. Plots will be created using the principle components overlaid with self-reported ethnic groups to aid in visualization of the data clustering and to determine the number of principal components required to infer genetic ancestry [Campbell, 2005]. The plots will be visually inspected to determine if modification of sub-groups, as outlined above, is warranted.

10.5. Evaluation of Genotypic Association

Logistic regression will be used to examine the effect of marker genotypes on the case-control status for diarrhoea (See Section 2.2 for diarrhoea case-control status definition). Prior to the PGx analysis, a model which includes necessary non-genetic independent variables that may influence the endpoints will be identified using the PGx Analysis Population (see Section 7.2 for the covariates that will be evaluated). The model may be reduced, if needed, to include only significantly predictive non-genetic terms ($p<0.05$) prior to beginning the genetic analyses. Multiple analyses will be run, evaluating each genetic marker, one at a time, while keeping covariates used constant throughout all analyses.

Proportional odds model, adjusting for some of the covariates, will be used to explore the effect of selected or all markers' genotypes on the ordinal diarrhoea toxicity grade levels if it is necessary. One hundred fifty one subjects classified as grade G1 diarrhoea toxicity level that are excluded in case-control analysis will be included in this analysis. The expected number of subjects in this analysis, by grade, is: 132 G0, 151 G1, and 90 G2-G4 (combined) subjects. The proportional odds assumption will be evaluated for the resulting three levels of diarrhoea toxicity. The analyses will be performed using ordinal logistic regression.

The following model will be fitted to identify the genetic effect.

Endpoint = PC1 + PC2 + … PC*n* + Covariates + Genotype -- model (1)

The endpoint variables are those listed in Section 7.1. Covariates are those listed in Section 7.2. PC1 represents the first principal component based on the analysis of genome-wide markers, PC2 represents the second principal component and so on. The number of principal components necessary to characterize subject ancestry, *n*, will be analytically determined; for many datasets an *n* of 2 or 3 is adequate. An additive genetic model is used here. *Genotype* codes for the number of copies (0, 1, or 2) of the minor (less frequent) allele and will be evaluated as a numeric variable.

10.5.1. Summary of Association Analysis Results

The p-value for the main effect of genotype for each marker determined to be statistically significant will be listed in table format. The distribution of genotypes and the HWE p-value for corresponding analysis population will also be listed.

Markers determined to be statistically significant according to the criteria outlined in Section 7.6 will also be summarized graphically. Bar plots displaying the genotypes against case/control status for the significant markers will be produced.

Regional plots that depict patterns of association within each gene as well as the linkage disequilibrium among the SNPs in the gene will be produced if there are markers that meet the Tier-specific significance criteria.

The relationship between association patterns and annotated gene features can provide substantial insight into the robustness of association results. These plots will be generated using the statistical package R and/or other software.

11. INTERPRETATION AND REPORTING OF ASSOCIATION ANALYSES RESULTS

11.1. Focus on Strongest Exploratory Results

The raw p-values will be reported. P-value thresholds for significant or suggestive significant associations are defined in Section 7.6. The secondary analysis will be performed only on markers that passed the p-value threshold for suggestive evidence in the primary analysis. Any genetic markers that meet the significant p-value thresholds will be investigated carefully for biological context via 1) whether has supportive evidence from the secondary analysis; and 2) phenotypic distribution for the genotypes whether supports a plausible genetic model and the association. Markers having p-values which don't meet the significant threshold but pass the suggestive significant threshold may also be investigated for phenotypic distribution for the genotypes supports a plausible genetic model and the association. Genetic markers that show significance or suggestive significance will be reported and assessed for further investigation.

Beyond assessment of the p-values, the criteria that will be used to characterize the strength of support for a genetic association may include some of the following:

- Are the genetic effects consistent with commonly observed genetic models (i.e. dominant, partial dominant, additive, or recessive)?
- Are there a sufficient number of subjects in the risk genotype group driving the association to yield strong statistical support?
- Is there further genetic or biological evidence to support the statistically significant associations with diarrhea?
- Are there other statistically significant SNPs in LD with the associated SNPs?

12. REFERENCES

Abraham B, et al. Drug-induced Diahrrea. *Current Gsstroenterology Reports.* 2007;9:365.

Campbell CD, et al. Demonstrating stratification in a European American population. *Nature Genetics.* 2005;37:868.

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

Stein A, et al. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. *Ther Adv Med Oncol.* 2010;2:51.

Sternberg CN, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol.* 2010;28:1061.

GlaxoSmithKline Document Number HJ2009/00002/00 Study ID VEG102616 and VEG105192. Whole Genome Scan Pharmacogenetics Reporting and Analysis Plan for the Investigation of Associations between Genetic Markers and Elevated Alanine Aminotransferase Levels and/or Total Bilirubin Concentration in subjects treated with Pazopanib (GW786034) in Studies VEG102616 and VEG105192. Report Date Mar-2009.

GlaxoSmithKline Document Number RM2008/00332/00 Study ID VEG102616 and VEG105192. PGx Reporting and Analysis Plan for Investigating the Relationship between Genetic Markers and ALT and/or Total Bilirubin Concentration in Subjects Treated with Pazopanib (GW786034) in Studies VEG102616 and VEG105192. Report Date 12-Aug-2008.

13. APPENDIX

13.1. Diarrhoea PGx Candidate Gene List

Table 13-1 Diarrhoea PGx Candidate Gene List

Gene group	Gene Name	Gene region (NCBI36)	Functional Marker	Numbers of functional markers (Tier 1)	Numbers SNPs/gene (Tier 1 and 2)
Pazopanib's mode of action	FLT1	Chr13: 27,774,389-27,967,265	rs111458691	1	98
	KDR	Chr4: 55,639,406-55,686,519	rs2071559 rs2305948 rs1870377 rs34231037	4	41
	FLT4	Chr5: 179,961,112-180,009,230	rs307826	1	41
	PDGFRA	chr4:54,790,021-54,859,169	rs1800812	1	21
	PDGFRB	chr5:149,473,595-149,515,615			51
	KIT	chr4:55,218,852-55,301,638			55
	VEGFA	chr6:43,845,931-43,862,201	rs699947 rs833061 rs1570360 rs2010963 rs3025039	5	18
Pazopanib ADME	CYP3A4	chr7:99,192,540-99,219,744	rs2740574	1	38
	CYP3A5	chr7:99083749-99115557	rs776746	1	42
	CYP2C8	chr10:96,786,519-96,819,244	rs10509681 rs11572103 rs1058930	3	62
	NR1I2	chr3:120,984,247-121,020,022	rs3814055	1	83
	ABCB1	chr7:86,970,884-87,180,500	rs1128503 rs2032582 rs1045642	3	204
	ABCG2	Chr4:89,230,440-89,299,035	rs2231137 rs2231142 rs72552713	3	67

Table 13-2 Diarrhoea PGx Candidate Gene List (continued)

Diarrhoea related genes	SLC6A4	chr17:25,549,032-25,586,841	rs25531 rs1042173	2	17
	IL10	chr1:205,007,571-205,012,462	rs1800896 rs1800872 rs1800871	3	14
	CFTR	chr7:116,907,253-117,095,954	rs73717525	1	50
	CLCA1	chr1:86,706,983-86,738,562			33
	SLC12A2	chr5:127,447,382-127,553,279			23
	ATP1A1	chr1:116,717,359-116,748,919			17
	ADCY3 ADCY6	chr2:24,895,542-24,995,559 chr12:47,446,242-47,464,144	 rs3730071	 1	36 9
	NOS3	chr7:150,319,080-150,342,609	RS2070744 RS1799983	2	21
	SLC9A3	chr5:526,425-577,447			31
	SLC5A1	chr22:30,769,259-30,836,645			27
	SLC26A3	chr7:107,193,148-107,230,914			22
	SCNN1A SCNN1B SCNN1G	chr12:6,326,274-6,354,976 chr16:23,221,092-23,300,121 chr16:23,101,541-23,135,701	rs2228576 rs3759324	2	15 27 24
	ANO1	Chr11:69,602,294-69,713,282			57
	TNF	chr6:31,651,329-31,654,091	rs1800629 rs361525	2	28
	IFNG	chr12:66,834,817-66,839,788	rs2069727 rs2069707	2	12
	IL1B	chr2:113,303,808-113,310,827	RS16944 rs1143627	2	11
	IL4	chr5:132,037,272-132,046,267	rs2243250 rs2070874	2	18
	IL6	chr7:22,733,323-22,738,145	RS1800795 RS1800796 RS10499563	3	24
	IL8	chr4:74,825,139-74,828,297	rs4073 rs1126647	2	7
	IL12A IL12B	chr3:161,189,323-161,196,500 chr5:158,674,369-158,690,059	rs3212227 rs568408	2	12 12
	IL13	chr5:132,021,764-132,024,700	rs1881457 rs1800925 rs20541	3	12
Total				53	1380