Title	:	Genetics Reporting and Analysis Plan for PGx7607: Investigation of germline genetic association with pyrexia by meta-analysis of melanoma subjects from BRF113710, BRF113929, BRF113683, MEK115306 and MEK116513.
Compound Number	:	GSK1120212+GSK2118436
Effective Date	:	24-OCT-2014

Description :

- This Genetics Reporting and Analysis Plan describes the planned analyses for eTrack study ID 202050 (PGx7607), an exploratory study to investigate associations between genetic variants and pyrexia in melanoma subjects. This study will use clinical and genetic data from subjects treated with dabrafenib or a combination of dabrafenib and trametinib from five clinical studies.
- The study objective is to identify germline genetic associations with pyrexia by meta-analysis of subjects treated with dabrafenib or a combination of dabrafenib and trametinib from studies BRF113710, BRF113929, BRF113683, MEK115306 and MEK116513

Subject :

Meta analysis, dabrafenib, combination of dabrafenib and trametinib, pharmacogenetics, PGx, genetics, safety, pyrexia, melanoma, 202050, PGx7607

Author's Name and Functional Area:

[redacted] Scientific Investigator (Statistical Genetics)	23-OCT-2014
[redacted] Therapy Area Specialist (Genetics)	23-OCT-2014
[redacted] Genetics Scientist (Genetics)	23-OCT-2014
[redacted] Head, Genetics Data Sciences (Genetics)	23-OCT-2014

Approved by:

[redacted] Senior Scientific Investigator (Statistical Genetics)	24-OCT-2014
[redacted] SM Director, Therapy Area Head (Genetics)	23-OCT-2014

TABLE OF CONTENTS

PAGE

1. GENETIC REPORTING & ANALYSIS PLAN SUMMARY

RAP Area	Description
Purpose	Planned analyses for eTrack study ID 202050 (PGx7607), which is an exploratory study to determine if genetic polymorphisms may predict development of pyrexia in patients receiving dabrafenib.
Primary Objective / Endpoint	Identify germline genetic associations with pyrexia by meta-analysis of subjects treated with dabrafenib or a combination of dabrafenib and trametinib from studies BRF113710, BRF113929, BRF113683, MEK115306 and MEK116513/Pyrexia
Study Design	Retrospective non-interventional pharmacogenetic study
Planned Analyses	Test for association between genetic variants and pyrexia, in subjects treated with dabrafenib or a combination of dabrafenib and trametinib from five clinical studies.
Analysis Populations	Caucasian metastatic melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib
Hypothesis	Germline genetic variants may be associated with pyrexia in dabrafenib-treated subjects
Primary Analyses	Test association between genetic variants and pyrexia in Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib using logistic regression
Secondary Analyses	• Test association between genetic variants and early onset pyrexia in Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib using logistic regression • Test association between genetic variants and time-to-pyrexia onset in Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib using Cox regression

2. SUMMARY OF KEY INFORMATION

2.1. Introduction and Rationale

Dabrafenib (GSK2118436) is a potent, ATP-competitive and selective inhibitor of mutant BRAF kinase (V600E/K) and trametinib (GSK1120212) is a selective, non-ATP competitive, allosteric inhibitor of MEK1 and MEK2 kinases. The U.S. Food and Drug Administration recently approved dabrafenib and trametinib as single-agent therapies as well as in combination for the treatment of unresectable melanoma or metastatic melanoma in adult patients with the most common type of BRAF mutations: BRAF V600E (dabrafenib) and BRAF V600E/K (trametinib). The BRAF V600E/K mutation is found in 40-60% of melanomas causing constitutive activation of BRAF and, in turn, the MAP kinase pathway.

Pyrexia, or fever, is one of the most common adverse events (AE) in subjects exposed to dabrafenib or a combination of dabrafenib and trametinib. The incidence of pyrexia is much higher (up to 70%) in subjects treated with a combination of dabrafenib and trametinib. The majority of these AEs are transient and resolve after treatment interruption, while a small proportion (2-5%) of subjects develops serious non-infectious febrile events such as influenza-like illness, cytokine release syndrome, and systemic inflammatory response syndrome which may require extensive management. The underlying mechanism for development of pyrexia on treatment with dabrafenib alone or in combination with trametinib is not clear. Prior pharmacogenetics (PGx) investigations of pyrexia (BRF116604/PGx6039 and 200997/PGx7550) in melanoma studies of dabrafenib or a combination of dabrafenib and trametinib (BRF113710, BRF113929, BRF113683 and MEK115306) identified no significant associations between pyrexia and potentially functional candidate, or genome wide variants or variants from HLA genes. However, with small sample sizes, the power to identify small or moderate genetic effects in these studies was limited.

This study aims to explore genome wide genetic associations by meta-analysis of melanoma subjects from BRF113710, BRF113929, BRF113683, MEK115306[1] and MEK116513[2]. In this study with the addition of subjects (n=172) from MEK116513, there will be more than an additional 10% power to detect genetic effects of common variants (MAF: 10-50%) compared to the previous investigation (200997/PGx7550).

[1] MEK115306 is a two arm phase III study comparing dabrafenib monotherapy with a combination of dabrafenib and trametinib – the 2 arms will be referred to as MEK115306-mono and MEK115306-combi, respectively.

[2] MEK116513 is a two arm phase III study comparing vemurafenib monotherapy (Roche BRAF inhibitor) with a combination of dabrafenib and trametinib – the combination arm will be referred to as MEK116513-combi.

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

Objectives	Endpoints
Primary	
Identify germline genetic associations with pyrexia by meta-analysis of Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib from studies BRF113710, BRF113929, BRF113683, MEK115306 (mono and combi) and MEK116513-combi.	Pyrexia
Secondary	
Identify germline genetic associations with early onset pyrexia (pyrexia developing within the first 8 weeks of treatment) by meta-analysis of Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib from studies BRF113710, BRF113929, BRF113683, MEK115306 (mono and combi) and MEK116513-combi.	Early onset pyrexia
Identify germline genetic associations with time-to-pyrexia onset by meta-analysis of subjects treated with dabrafenib or a combination of dabrafenib and trametinib from studies BRF113710, BRF113929, BRF113683, MEK115306 (mono and combi) and MEK116513-combi.	Time-to-pyrexia onset

Endpoint definitions

- **Pyrexia** (Case-control study)
    - **Case** is defined as any metastatic melanoma subject with normal temperature at baseline (< 38 ºC) and developing an AE of pyrexia (grade ≥2* according to *NCI Common Terminology Criteria for Adverse Events (CTCAE) v.4*)[3] while receiving treatment.

    * Subjects with grade 1 (38-39ºC) fever are excluded from the definition of case because of the possible non-specificity of low-grade fever (related to study drug vs. due to other underlying conditions).

[3] http://evs.nci nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

Endpoint definitions

- **Control** is defined as a metastatic melanoma subject who received dabrafenib or a combination of dabrafenib and trametinib for the days in Table 1 (which corresponds to the time by which 90% of the cases have had an AE of pyrexia)**, had normal temperature (< 38.0 ºC) at baseline, and no pyrexia AEs throughout the treatment duration.

** To ensure that the subjects in the control population had sufficient cumulative exposure to dabrafenib or its combination with trametinib and to reduce the risk of including 'hidden' cases (subjects that could have become pyrexia cases had they been exposed to dabrafenib longer), a conservative selection of controls was needed. The 90% percentile for pyrexia onset was adopted from BRF116604/PGx6039 and 200997/PGx7550 results which were discussed with BRAF/MEK project team.

- **Early onset pyrexia (case-control study)**
  - Case is defined as a metastatic melanoma subject who developed an AE of pyrexia (grade 2 or higher) on or before 56 days (8 weeks)†.
  - Control is defined as same as in the pyrexia above.

† The cut-off of 8 weeks for early onset of pyrexia was decided based on the observation that more than 60% of the pyrexia cases across all the 5 melanoma studies under investigation had had an event of grade ≥ 2 pyrexia within the first 8 weeks of treatment. This was statistically significant when compared to the remaining cases that had onset of pyrexia ranging from 60-400 days (up to 57 weeks) of treatment.

- **Time-to-pyrexia onset** (continuous study)
  - Time-to-pyrexia onset is defined as the number of study days from initiation of treatment until the first grade ≥2 pyrexia event. Subjects who do not have pyrexia will be censored at the end of total cumulative days of study treatment.

Table 1 Number of control subjects who have cumulative duration of exposure ≥ time to pyrexia onset in cases across different clinical studies

		Time to pyrexia onset in days (for cases)				The number of controls who have cumulative duration of exposure ≥ time to pyrexia onset in cases			
	Percentile	95%	90%	75%	50%	95%	90%	75%	50%
Clinical study	BRF113710, BRF113929 and BRF113683	178	139	82	23	76	130	210	243
	MEK115306-mono	271	142	67	27	50	89	105	117
	MEK115306-combi	208	182	114	41	55	56	66	76
	MEK116513-combi	257	190	93	41	76	86	114	120

2.3. Study Design

Overview of Key Study Design Features	
Brief Design Features	• A brief description of each study and PGx samples are provided in Table 2 and Table 3 below, respectively. • Meta-analysis will be conducted on Caucasian melanoma subjects from five metastatic melanoma studies (BRF113710, BRF113929, BRF113683, MEK115306, and MEK116513). The subjects from BRF113710, BRF113929, BRF113683 and MEK115306 were analyzed in a prior PGx investigation of pyrexia, 200997/PGx7550 (see RAP and results for details). The dabrafenib and trametinib combination arm in MEK116513 (MEK116513-combi) will be analyzed independently and meta-analyzed with results from 200997/PGx7550.

Table 2 Description of five clinical studies

Clinical Study Number	Brief Description of Study	Indication	Phase
BRF113710	A Phase II single-arm, open-label study of dabrafenib in BRAF-mutant metastatic melanoma.	BRAF-mutant metastatic melanoma	II
BRF113929	A Phase II open-label, two-cohort, multicentre study of dabrafenib as a single agent in treatment naïve and previously treated subjects with BRAF mutation-positive metastatic melanoma to the brain (BREAK-MB).	BRAF mutation-positive metastatic melanoma to the brain	II
BRF113683	A Phase III randomized, open-label study comparing dabrafenib to DTIC in previously untreated subjects with BRAF mutation positive advanced (Stage III) or metastatic (Stage IV) melanoma.	BRAF mutation positive advanced or metastatic melanoma	III
MEK115306	A two-arm, randomized, double-blinded, multi-center Phase III study to evaluate efficacy and safety of dabrafenib + trametinib compared to dabrafenib + trametinib-placebo in subjects with unresectable (Stage IIIC) or metastatic (Stage IV) melanoma.	BRAF mutation-positive melanoma	III
MEK116513	A Phase III, randomized, open-label study comparing the combination of the BRAF inhibitor, GSK2118436 and the MEK inhibitor, GSK1120212 to the BRAF inhibitor vemurafenib in subjects with advanced (Stage IIIc) or metastatic (Stage IV) BRAF V600E/K mutation-positive cutaneous melanoma	BRAF mutation-positive melanoma	III

Table 3 Summary of PGx samples

Study	BEF113710 + BRF113929 + BRF113683	BEF113683	MEK115306		MEK116531	Total	Cases	Controls
	# treated with dabrafenib	# crossed over from DTIC to dabrafenib	Dabrafenib only (mono)	Darafenib + Trametinib (combi)	Darafenib+ Trametinib (combi)			
PGx population	342	23	177	184	289	1015		
Missing	6					6		
No fever	243	17	118	79	123	580		361
Grade 1	50	2	32	47	79	210		
Grade 2	39	3	23	46	68	179	179	
Grade 3	4	1	4	12	18	36	36	
Total							**218**	**361**

2.4. Statistical Hypotheses

Germline genetic variants may be associated with pyrexia in melanoma subjects treated with dabrafenib or a combination of darafenib and trametinib.

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

The clinical studies being examined were not prospectively designed to address PGx research hypotheses and, thus, may not have statistical power to detect moderate genetic effects. Post-hoc assessment of statistical power is necessary to determine the sizes of effects that can be detected given the admissible PGx data. The distribution of genotypes varies considerably from one genetic marker to the next (i.e., the genotype data within each genetic marker will not be balanced), so the statistical power of each analysis cannot be guaranteed. However, power can be estimated by assuming a range of possible risk allele frequencies and genetic effect sizes in meta-analysis.

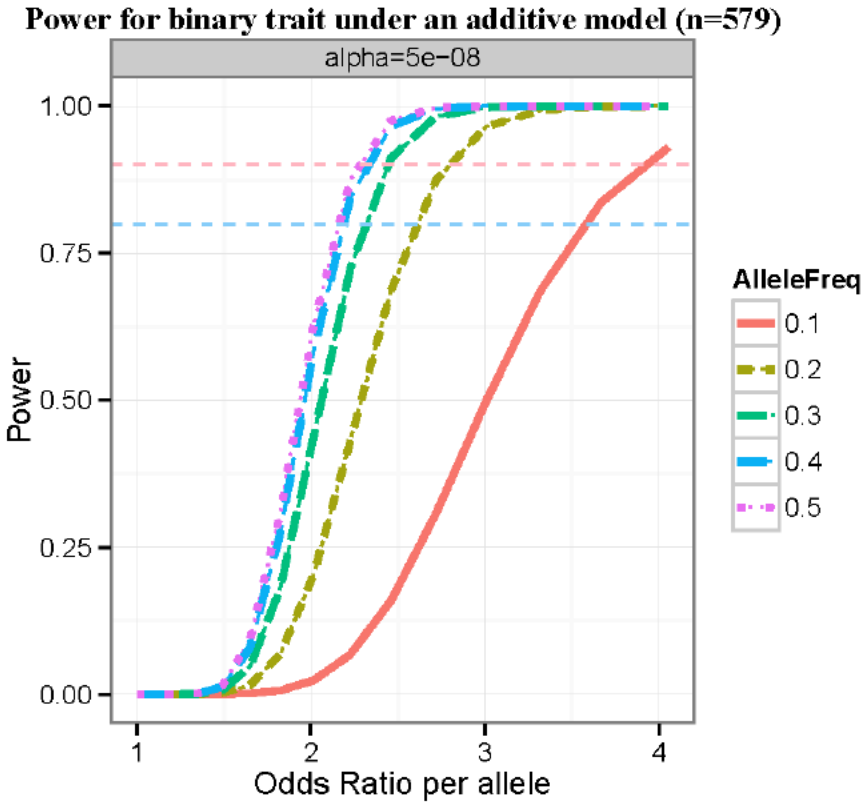
Statistical power to detect a genetic effect using all 218 pyrexia cases and 361 controls was evaluated assuming an additive genetic model with varying allele frequencies (5-50%), and the GWAS threshold for declaring statistical significance ($p=5x10^{-8}$) to control for multiple testing. The statistical power estimates are plotted (Figure 1). The different colour curves represent varying power estimates for a range of allele frequencies. The light pink and light blue dashed lines represent 90% and 80% powers, respectively.

Given 218 cases and 361 controls, less common genetic variants (MAF ≤ 10%) may only have 80% power to detect moderately large genetic effects (OR per allele > 3.6).

However, with relatively common variants (MAF > 10%), the current study has 80% power to detect smaller genetic effects (OR per allele > 2.2), assuming an additive genetic model.

There was 80% power for GWAS in 200997/PGx7550 to detect moderate effect sizes (OR per allele > 2.5) for MAF > 10%. The power to detect moderate effect sizes has improved in this study with the addition of subjects from MEK116513-combi. But, we acknowledge that the power is limited to detect GWAS level effect sizes of OR < 2.2 or variants with MAF < 20% on a genome wide scale.

Figure 1 Power for CC analysis. The power is calculated assuming 218 cases and 361 controls



4. PG$_X$ ANALYSIS POPULATIONS

Population	Definition / Criteria	Endpoint(s) Evaluated
Primary case/control	• Caucasian melanoma subjects treated with dabrafenib or a combination of dabrafenib and trametinib: will comprise all subjects who provided written informed consent for PGx research, provided a blood sample for genotyping and were successfully genotyped	• Pyrexia

Population	Definition / Criteria	Endpoint(s) Evaluated
	for at least one of the genetic variants under study, have valid phenotype data and passed genotyping QC. • This analysis population will include the metastatic melanoma subjects who met the definition of a case or control (see Section 2.2). • For genetic association analyses, the groups of subjects analyzed in the previous meta-analysis (200997/PGx7550) will be kept as before and subjects from the combination arm of MEK116513 will be added to the meta-analysis as follows: ○ BRF113710+BRF113929+BRF113683 ○ MEK115306-mono ○ MEK115306-combi ○ MEK116513-combi	
Secondary (Early onset)	• This is a subpopulation of the primary case/control population. • Cases will be subjects who developed an AE of pyrexia (grade 2 or higher) on or before 56 days (8 weeks) (see Table 4 for summary population). Controls will be same as in the primary case/control population (as described in Section 2.2).	• Early onset pyrexia
Secondary (Time-to-pyrexia onset)	• This population comprises all subjects in the primary case/control population. • Events will be melanoma subjects with grade ≥2 pyrexia. The time to onset will be cumulative days of study treatment until the first event of grade ≥2 pyrexia. • Non-events (censored) will be subjects who do not have pyrexia. The censoring time will be the total cumulative days of study treatment.	• Time-to-pyrexia onset

Table 4 Summary of population for secondary PGx analysis of early onset pyrexia

Study	BEF113710+ BRF113929+ BRF113683	MEK115306		MEK116531	Total
	# treated with dabrafenib+ crossed over from DTIC to dabrafenib	Dabrafenib only (mono)	Darafenib+ Trametinib (combi)	Darafenib+ Trametinib (combi)	
# early onset cases/# cases (%)	33/47 (70)	18/27 (67)	35/58 (60)	50/86 (58)	136/218 (63)
Controls	130	89	56	86	361

5. CONSIDERATIONS FOR DATA ANALYSES

Genetic Variants	• Genome-wide variants ○ Two different platforms were used to genotype subjects from different studies: Illumina OmniExpressExome for BRF113710, BRF113929 and BRF113683 and Affymetrix Axiom Biobank Plus GSK Custom array for MEK115306 and MEK116513. In order to have a common set of variants across the subjects from 5 different clinical studies for meta-analysis, the variants for MEK116513 subjects will be imputed from the 1000 Genomes Project whole genome sequence data using an in-house software pipeline (see APPENDIX 8). The variants for subjects from the other 4 clinical studies were previously imputed in 200997/PGx7550 using the same pipeline. ○ The variants from the HLA region will also be imputed to 4-digit resolution in pyrexia cases and controls using HIBAG (see APPENDIX 8). These variants will be tested for associations along with the genome wide variants.
Variant Category and Type I Error	• GWAS: False positives will be controlled at 5% per analysis. • HLA variants: False positives will be controlled at 5% per analysis.

6. DATA HANDLING CONVENTIONS

Details of data handling conventions are provided in the APPENDIX 1 to APPENDIX 3.

7. PHARMACOGENETIC ANALYSES

7.1. Primary Analyses

+++ Primary Statistical Analysis +++
Endpoint / Covariates/Model Specification
Pyrexia: Logistic regression, adjusting for sex and genetic ancestry estimates
Analysis Population
Primary case/control
Genetic Variants
• GWAS • HLA
Effects to be Modeled (Main or Interaction Effect; Dominant/Additive/Recessive Genetic Model)
Genetic main effects, assuming an additive genetic model (i.e, genetic variants will be coded as 0/1/2 representing **the number of copies of the minor allele at each locus)**
Meta-analysis
• The genetic association study with pre-specified covariates will be conducted in subjects from MEK116513-combi arm first and then the results will be meta-analyzed along with summary statistics from the other 4 clinical studies (BRF113710, BRF113929, BRF113683 and MEK115306 (mono and combi arms)) which were analyzed in the previous PGx investigation (200997/PGx7550). • A random effect inverse variance weighted meta-analysis of the effect size estimates will be conducted.
Statement Regarding What Constitutes a Significant Result
• GWAS: $P< 5x10^{-8}$ for GWAS • HLA markers: a Bonferroni correction for the effective number of tests after accounting for the correlation among number of variants looked up [Moskvina, 2008] will be used.

7.2. Secondary Analyses

+++ Secondary Statistical Analysis 1 +++
Endpoint / Covariates/Model Specification
Early onset pyrexia: Logistic regression, adjusting for sex and genetic ancestry estimates
Analysis Population
Secondary (Early onset)
Genetic Variants
• GWAS

+++ Secondary Statistical Analysis 1 +++
• HLA
Effects to be Modeled (Main or Interaction Effect; Dominant/Additive/Recessive Genetic Model)
Same as in the primary analysis.
Meta-analysis
Same as in the primary analysis.
Statement Regarding What Constitutes a Significant Result
Same as in the primary analysis.
+++ Secondary Statistical Analysis 2 +++
Endpoint / Covariates/Model Specification
Time-to-pyrexia onset: Cox regression, adjusting for sex and genetic ancestry estimates.
Analysis Population
Secondary (Time-to-pyrexia onset)
Genetic Variants
• GWAS • HLA
Effects to be Modeled (Main or Interaction Effect; Dominant/Additive/Recessive Genetic Model)
Same as in the primary analysis.
Meta-analysis
• The association analysis with the pre-specified covariates will be conducted separately within each indication group (see analysis population definition). • Meta-analysis will be done across groups for each genetic variant using a random effect inverse variance weighted method.
Statement Regarding What Constitutes a Significant Result
Same as in the primary analysis.

7.3. General Pharmacogenetic Analysis Conventions

General pharmacogenetic analysis conventions are provided in APPENDIX 4 to APPENDIX 8.

8. REFERENCES

Buhm Han and Eleazar Eskin (2011). Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *The American Journal of Human Genetics*, 88:586-598.

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

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.

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.

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.

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

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). Manhattan plots and Quantile-Quantile (QQ) plots may be used to visualize P-values at the whole genome scale. Results may be annotated by whether the genetic variant was typed or imputed, and a metric for quality of imputation. 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:Premature Withdrawals & Handling of Missing Data

9.2.1. 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.

9.2.2. 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.

9.3. APPENDIX 3:Genotype/Subject Quality Control

9.3.1. 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 9.3.2.

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.

9.3.2. Genotype Quality Control

Prior to genotype imputation (see APPENDIX 8), 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] .

9.4. APPENDIX 4:Multiple Comparisons & Multiplicity

9.4.1. GWAS analysis

The conventional $P \leq 5x10^{-8}$ threshold for declaring genome-wide significance for common variants (MAF>=5%) will be used [McCarthy, 2008] [Dudbridge, 2008].

For the "candidate variant lookup" analyses within the GWAS, a Bonferroni correction for the effective number of tests after accounting for the correlation among number of variants looked up [Moskvina, 2008] will be used.

9.5. APPENDIX 5: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).

9.6. APPENDIX 6: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 (measured as D'/r^2) 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 of interest.

9.7. APPENDIX 7: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, 2006; Patterson, 2006; Novembre, 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.

9.8. APPENDIX 8:Genotype Imputation

Genotype imputation for genetic variants that were not genotyped on Affymetrix Atom ("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].

9.9. APPENDIX 9:Abbreviations & Trade Marks

9.9.1. Abbreviations

Abbreviation	Description
AE	Adverse Event
CC	Case-Control
DNA	Deoxyribonucleic acid
Gx RAP	Genetics Reporting & Analysis Plan
GSK	GlaxoSmithKline
HWE	Hardy-Weinberg Equilibrium
LD	Linkage Disequilibrium
PC	Principal Component
PGx	Pharmacogenetics
RAP	Reporting & Analysis Plan
SE	Standard Error
SNP	Single Nucleotide Polymorphism

9.9.2. Trademarks

Trademarks of the GlaxoSmithKline Group of Companies
MEKINIST
TAFINLAR

Trademarks not owned by the GlaxoSmithKline Group of Companies
None