Open

npg

## ORIGINAL ARTICLE

# Sequencing of Lp-PLA2-encoding *PLA2G7* gene in 2000 Europeans reveals several rare loss-of-function mutations

K Song[1], MR Nelson[2], J Aponte[2], ES Manas[4], S-A Bacanu[2], X Yuan[1], X Kong[1], L Cardon[1], VE Mooser[1], JC Whittaker[3] and DM Waterworth[1]

[1]*Department of Genetics, GlaxoSmithKline, Upper Merion, PA, USA;* [2]*Department of Genetics, Research Triangle Park, Durham, NC, USA;* [3]*Department of Genetics, Harlow, UK and* [4]*Department of Computational and Structural Chemistry, GlaxoSmithKline, Upper Merion, PA, USA*

**Correspondence:**
Dr DM Waterworth, Department of Genetics, GlaxoSmithKline, Upper Merion, PA 19406, USA.
E-mail: dawn.m.waterworth@gsk.com

Elevated plasma levels of lipoprotein-associated phospholipase $A_2$ (Lp-PLA2) activity have been shown to be associated with increased risk of coronary heart disease and an inhibitor of this enzyme is under development for the treatment of that condition. A Val279Phe null allele in this gene, that may influence patient eligibility for treatment, is relatively common in East Asians but has not been observed in Europeans. We investigated the existence and functional effects of low frequency alleles in a Western European population by re-sequencing the exons of *PLA2G7* in 2000 samples. In all, 19 non-synonymous single-nucleotide polymorphisms (nsSNPs) were found, 14 in fewer than four subjects (minor allele frequency < 0.1%). Lp-PLA2 activity was significantly lower in rare nsSNP carriers compared with non-carriers ($167.8 \pm 63.2$ vs $204.6 \pm 41.8$, $P = 0.01$) and seven variants had enzyme activities consistent with a null allele. The cumulative frequency of these null alleles was 0.25%, so < 1 in 10 000 Europeans would be expected to be homozygous, and thus not potentially benefit from treatment with an Lp-PLA2 inhibitor.
*The Pharmacogenomics Journal* (2012) **12,** 425–431; doi:10.1038/tpj.2011.20; published online 24 May 2011

## Introduction

Lipoprotein-associated phospholipase $A_2$ (Lp-PLA2), encoded by *PLA2G7*, is abundant in coronary atherosclerotic plaque and generates pro-inflammatory mediators from oxidized low-density lipoprotein particles. Elevated levels of the enzyme activity have been associated with an increased risk of coronary heart disease in large epidemiological studies.[1] These observations have prompted the development of inhibitors (darapladib) to this enzyme as a treatment for coronary heart disease.[2]

One variant within *PLA2G7*, Val279Phe, ranges in allele frequency from 4 to 18% in East Asian populations and abolishes enzyme activity,[3] so that 279Val/279Phe heterozygotes have roughly 50% of the activity of homozygous 279Val carriers, and homozygous carriers of the minor allele have almost no measurable activity. This variant has been only rarely observed within Europeans, though other common variants with very small effects on enzyme activity have been identified in various ethnic groups.[4] Here, we hypothesized that rare coding variants with large effects on enzyme activity would exist in Europeans, as is the case for other traits, such as triglycerides[5] and blood pressure[6] and that these

variants may account for a certain proportion of the inter-individual variability in plasma Lp-PLA2 levels.

To test this hypothesis, we re-sequenced the *PLA2G7* coding regions of 2000 individuals from the CoLaus population-based study in Lausanne, Switzerland[7] to identify coding single-nucleotide polymorphisms (SNPs) and more specifically non-synonymous SNPs (nsSNPs). We then predicted the effect of these nsSNPs on protein function by means of Sorting Intolerant from Tolerant (SIFT[8,9]) and Polymorphism Phenotyping (PolyPhen[10,11]) algorithms and estimated their effect on enzyme activity. We also performed a qualitative structural assessment based on the available crystal structure of Lp-PLA2 enzyme, to assess the degree of corroboration with both the SIFT and PolyPhen algorithms. Subsequently, we examined the association between nsSNPs in *PLA2G7* and a variety of cardiovascular, metabolic and psychiatric traits to determine the phenotypic consequences of carriage of likely loss-of-function alleles.

## Materials and methods

### Study sample
CoLaus is a population-based study of 6188 Lausanne residents aged 35–75 years.[7] In brief, the study assessed cardiovascular disease risk factors and collected DNA, plasma, serum and urine samples. For the current study, we limited our evaluation to 2000 unrelated CoLaus subjects aged 35–66 years who had also undergone a psychiatric evaluation (PsyCoLaus[12]). The project was approved by the Local Ethics Committee and each participant consented for the use of their data and samples for this study.

### DNA sequencing
Assay design for directed sequencing was performed at Beckman Coulter Genomics (Danvers, MA, USA) using the Beckman Coulter Genomic amplicon design software based on NCBI RefSeq build 36.3. PCR was performed using validated primer sets and standard conditions. Sequencing was performed on both strands using the M13 forward and reverse priming sites in a 384-well format. Sequencing reagents were subjected to thermal cycling at $95\,^{\circ}C$ for $10\,s$, $50\,^{\circ}C$ for $5\,s$ and $60\,^{\circ}C$ for $2\frac{1}{2}\,min$ for 40 cycles. The PCR products were loaded on ABI 3730xl automated sequencing instruments (Foster City, CA, USA). Amplicons that passed quality review were analyzed to identify single base differences relative to the reference sequence using Poly-Phred v.5.04 and v.6.0.[13,14] Genotype calls for all subjects at each coding variant position were manually reviewed.

### Enzyme activity measurement
Lp-PLA2 activity was measured on $300\,\mu l$ of EDTA-sourced plasma samples using a colorimetric activity method at Quest Diagnostics (Nicholls Institute). Plasma samples were reacted with 1-myristoyl-2-(4-nitrophenylsuccinyl) phosphatidylcholine in a 96-well microtiter plate. The change in the absorption at $405\,nm$ in kinetic mode, which is proportional to the activity of Lp-PLA2 in the sample, was measured. The serum samples had been stored for 3.4–6.4 years.

### Statistical analyses
Association between plasma Lp-PLA2 activity and prediction score by SIFT or PolyPhen was assessed using a robust linear regression model adjusting for storage duration of serum, carrier status and gender. Multiple logistic and linear regression was performed with selected covariates to test the association between common variants (defined here as minor allele frequency (MAF) $> 0.5\%$) and clinical phenotypes. Rare variants (MAF $< 0.5\%$), were analyzed by aggregating across variants.[15] Two analyses were done, the first aggregating all non-synonymous rare variants and the second including only those predicted to be deleterious by PolyPhen and SIFT. Covariates included age, gender, smoking status, alcohol usage, physical activity and the first five principle components, computed from previously reported genome-wide genotype data[16,17] using EIGENSOFT version 2.0.[18] All statistical analyses were carried out using SAS v9.2 (Cary, NC, USA). A 5% significance threshold was determined by adjusting for the number of analyzed traits using Bonferroni's correction. We estimated the proportion of variance in Lp-PLA2 activity explained by rare variants using standard quantitative genetic formulae[19] by considering aggregated rare variants as a single locus, assuming an additive model and using observed genotypic values and MAF. The genotypic values for this calculation were estimated by multivariate regression adjusted for storage duration of serum, high-density lipoprotein cholesterol and low-density lipoprotein cholesterol.

## Results

We re-sequenced all coding (1325 bp) and non-coding (4909 bp) exons, as well as flanking intronic sequence (513 bp), of *PLA2G7* in 2000 unrelated CoLaus subjects. A total of 128 variants were identified (see Supplementary Figure 1), including 26 coding variants of which 19 resulted in amino-acid changes (Table 1). Most non-synonymous variants (16; 84%) were rare (MAF $< 0.5\%$), of which 10 (53%) were observed in just one heterozygous individual, that is, singleton variants. One individual was heterozygous for two rare nsSNPs, Arg82His (MAF $= 0.08\%$) and Asp181Gly (singleton) of unknown phase. We observed one carrier of the Val279Phe null allele commonly found in East Asians. Principal component analysis using previously generated genome-wide association study data showed that this individual clusters with other CoLaus subjects having Swiss rather than Asian ancestry (see Supplementary Figure 2), demonstrating that this allele is present at very low frequency in Europeans.

SIFT and PolyPhen protein function predictions were highly concordant. Both algorithms predicted the three common variants (Arg92His, Ile198Thr and Val379Ala) to be benign. Of the rare variants, 12 were predicted to be damaging and 3 benign (Table 1). Only one variant

**Table 1** Non-synonymous variants observed in *PLA2G7* and predicted functional effects

| Amino acid | MAF[a] | SIFT | PolyPhen | Structure-based assessment of impact on protein function |
|---|---|---|---|---|
| **Phe51Ile** | **0.0015** | **Tolerant** | **Benign** | **No** |
| **Asp69Gly** | **0.00025** | **Deleterious** | **Probably damaging** | **Yes** |
| **Arg82His** | **0.00076** | **Deleterious** | **Probably damaging** | **Yes** |
| Arg92His | 0.25 | Tolerant | Benign | No |
| Phe110Leu | 0.00025 | Deleterious | Probably damaging | Maybe |
| **Trp134\*\*\*[b]** | **0.00031** | **Deleterious** | **Probably damaging** | **Yes** |
| **Asp181Gly** | **0.00026** | **Deleterious** | **Probably damaging** | **Yes** |
| **Thr187Pro** | **0.00025** | **Deleterious** | **Probably damaging** | **Yes** |
| **Lys191Asn[c]** | **0.00025** | **Deleterious** | **Probably damaging** | **No** |
| Ile198Thr | 0.052 | Tolerant | Benign | No |
| **Asp200His[c]** | **0.00025** | **Deleterious** | **Benign** | **No** |
| **Ser273Phe** | **0.00025** | **Deleterious** | **Probably damaging** | **Yes** |
| Val279Phe[d] | 0.00025 | Deleterious | Probably damaging | Yes |
| **Leu283Pro** | **0.00050** | **Deleterious** | **Probably damaging** | **Yes** |
| **Gly303Cys** | **0.0018** | **Deleterious** | **Possibly damaging** | **Maybe** |
| **Ala326Val** | **0.00025** | **Tolerant** | **Benign** | **No** |
| **Met331Lys** | **0.00025** | **Deleterious** | **Probably damaging** | **Yes** |
| Val379Ala | 0.23 | Tolerant | Benign | No |
| **Glu414Asp** | **0.00035** | **Tolerant** | **Benign** | **No** |

Abbreviations: MAF, minor allele frequency; PolyPhen, Polymorphism Phenotyping; SIFT, Sorting Intolerant from Tolerant.
Bold type indicates the novel non-synonymous variants found in this study.
[a]MAF in 2000 Europeans.
[b]\*\*\*: stop.
[c]Disagreements among prediction methods.
[d]Asian null variant known from the literature.

(Asp200His) was predicted to be damaging by SIFT but benign by PolyPhen.

A qualitative assessment of the expected impact of these amino-acid changes was also performed based on analysis of an Lp-PLA2 crystal structure (Table 1). As shown in Figure 1, the SNPs are widely distributed throughout the three-dimensional structure. In cases where an SNP occupies a position within a particular secondary structural element and is buried within the core of the protein without any apparent room for that particular residue substitution, it was predicted that there could be an impact on protein stability and possibly function. Likewise, if the SNP lies within or near the active site and appears likely to directly, or indirectly affect protein function (for example, via substrate binding or turnover), then such a prediction was also made, with mutation of the catalytic serine being an obvious example. In contrast, residues with side chains that are highly solvent exposed and far from the active site were predicted to not have an effect on function. Some cases appeared borderline and were classified as such. Results were found to be generally concordant with the SIFT and PolyPhen predictions, and in agreement with a smaller scale assessment of the impact of polymorphisms by Samanta and Bahnson.[20] There was one exception though, in that Lys191Asn was predicted to have a negative impact on function by SIFT and PolyPhen, but not by the structural assessment. The basis of the structural prediction was that Lys191 is part of a solvent-exposed loop and is roughly 30 Å from the catalytic serine. However, it cannot be ruled out

that phenotypic effects due to this SNP might be due to modulation of lipoprotein binding interactions, as pointed out by Samanta and Bahnson for another nearby solvent-exposed SNP (Ile198Thr).

To determine the effect of these nsSNPs on Lp-PLA2 enzyme activity, we selected all 29 rare nsSNP carriers with a serum sample available for analysis. Variant carriers were matched for age (within 1 year), sex and low-density lipoprotein cholesterol (within 1 mmol l$^{-1}$) with non-carriers in a 1:2 ratio. Characteristics of the subjects selected for enzyme activity analysis are summarized in Table 2. Only mean of storage duration of serum sample showed a statistically significant difference ($P = 0.01$), with carrier samples having been stored for a 0.4-year shorter period. To account for the influence of serum storage duration assay results, it was included as a covariate in all subsequent analyses of enzyme activity.

The mean Lp-PLA2 enzyme activity was significantly lower in carriers than in non-carriers (Table 2). Individual carrier enzyme activities, adjusted for storage duration of serum and sex, are shown in Figure 2, expressed as standard deviations from the mean of non-carriers. The individual who carried the Val279Phe mutation was a white male, French-speaking resident of Lausanne reported to have all four grandparents born in Switzerland and demonstrated a genetic background similar to other French-speaking Lausanne residents (see Supplementary Figure 2).[21] He had an enzyme activity that was 38% of the mean of non-carriers (2.8 s.d. below), consistent with previous studies of the

Val279Phe variant. One subject with a non-sense mutation at position Trp134, which would certainly result in a non-functional protein, was 1.5 s.d. below the mean, consistent with the range of Lp-PLA2 activity expected for carriers of a single functional gene copy. The individual with lowest enzyme activity was the subject carrying both Arg82His and Asp181Gly. The measured activity for this subject was 21.1 nmol ml$^{-1}$ per min, 10% of the mean of non-carriers, a level approaching values typically observed in homozygous non-functional Val279Phe variant carriers,[3,22] suggesting that this subject is a compound heterozygote. This individual was a 39-year-old England-born man who is healthy with unremarkable phenotype and no evident disease. Seven carriers with the Gly303Cys variant had average activity of 173.2 nmol ml$^{-1}$ per min, 84% of the mean of non-carriers, though this reduction was not statistically significant ($P = 0.20$). Altogether, there were seven variants carried by nine subjects with a cumulative frequency of 0.25% with effects similar to those expected for a null variant with enzyme activity > 1 s.d. below the mean of non-carriers. The six Phe51Ile carriers, an allele predicted to be benign, had an average activity of 241.7 nmol ml$^{-1}$ per min, 0.9 s.d. above the mean ($P = 0.02$), suggesting that this variant may cause an increase in enzyme function. One other singleton variant, Phe110Leu, had a similarly increased activity. There was significant association between Lp-PLA2 enzyme activity and functional predictions obtained by both PolyPhen ($P < 0.0001$; Figures 2 and 3) and SIFT ($P = 0.0001$, data not shown). It can further be observed in Figure 1, that the residues of the likely null variants based on activity measurements are found in the core of the protein, close to the active site. Conversely, the non-null variants are predominantly on the surface of the protein.

We further investigated the effects of both common and rare nsSNPs in *PLA2G7* on other cardiovascular, metabolic and psychiatric traits. None of the three common variants—predicted to be benign—were significantly associated with any trait. Aggregate analysis of rare nsSNPs (cumulative frequency = 0.8%) also resulted in no statistically significant associations after Bonferroni's correction (see Supplementary Figure 3). Similar association tests were performed for the 13 predicted functional variants and for the 7 variants that lead to lower enzyme activity. Neither approach resulted in statistically significant associations (see Supplementary Figure 3).

## Discussion

Complete re-sequencing of the *PLA2G7* coding regions in 2000 Europeans identified seven rare non-synonymous changes in nine carriers, including one compound heterozygote, that lead to a substantial reduction in enzyme activity. The null allele found at MAFs up to 18% in East Asians (Val279Phe) was found in one European individual yielding a MAF estimate of 0.03% (95% confidence interval 0–0.07). This study yields an estimate of the cumulative frequency of null alleles in the European population to be 0.25% (95% confidence interval 0.09–0.41). Although the rare loss-of-function variants have relatively large effects on enzyme activity, their rarity means that they can explain only a small proportion ($\sim$3%) of the European variation. This compares with an estimate of < 1% for Val379Ala, which although common in Europeans, has a very small effect on enzyme activity.[23]
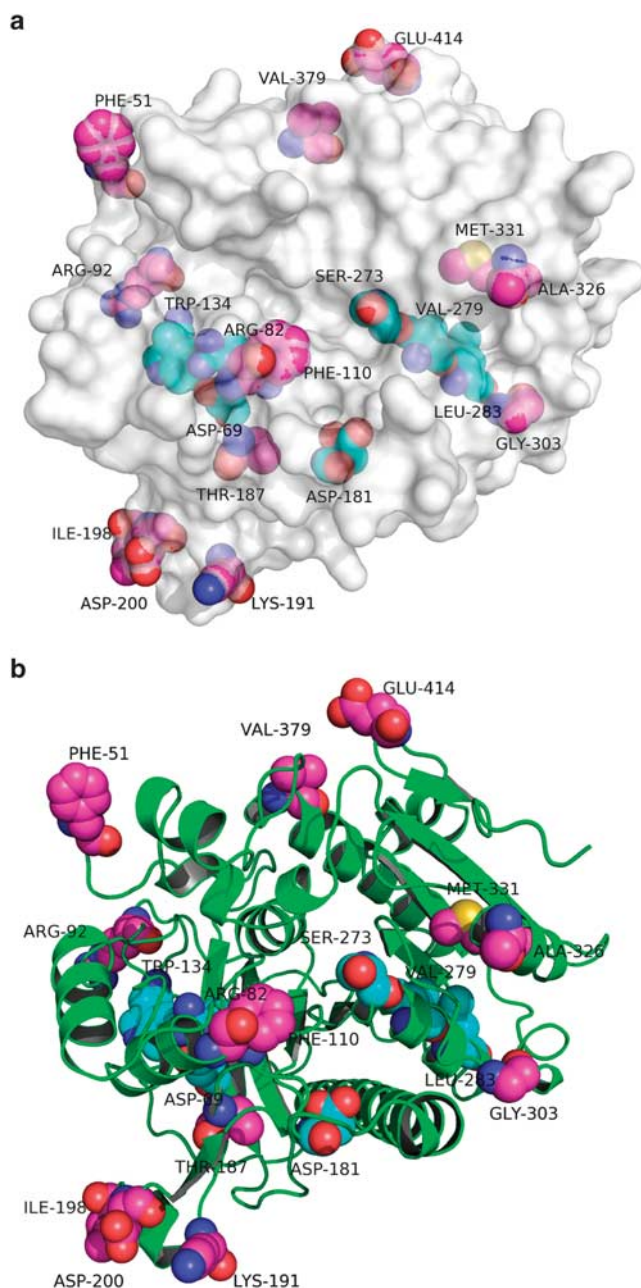


**Figure 1** Crystal structure of human lipoprotein-associated phospholipase A$_2$ (Lp-PLA2) showing (**a**) single-nucleotide polymorphisms (SNPs) rendered as space-filling spheres, along with a semitransparent molecular surface indicating the solvent exposure in a qualitative manner and (**b**) a cartoon representing the protein backbone, indicating where SNPs lie along different secondary structural elements.

**Table 2** Characteristics of carriers of *PLA2G7* rare non-synonymous variants and matched non-carriers

|  | Carriers (N = 29)[a] | Non-carriers (N = 58) | P-value |
|---|---|---|---|
| Female | 13 (45%) | 26 (45%) | *[b] |
| Age (years) | 50.1 ± 8.2 | 50.1 ± 8.2 | * |
| Body mass index (kg m$^{-2}$) | 25.9 ± 4.2 | 25.3 ± 4.1 | 0.53 |
| LDL cholesterol (mmol l$^{-1}$) | 3.6 ± 0.95 | 3.5 ± 0.80 | 0.54 |
| HDL cholesterol (mmol l$^{-1}$) | 1.5 ± 0.37 | 1.6 ± 0.42 | 0.07 |
| Previous or current alcohol use | 21 (72%) | 45 (76%) | 0.60 |
| Current smoking | 9 (31%) | 13 (22%) | 0.44 |
| Storage duration of serum (years) | 4.6 ± 0.67 | 5.0 ± 0.80 | 0.04 |
| Lp-PLA2 activity (nmol ml$^{-1}$ per min) | 167.8 ± 63.2 | 204.6 ± 41.8 | 0.01[c] |

Abbreviations: HDL, high-density lipoprotein; LDL, low-density lipoprotein; Lp-PLA2, lipoprotein-associated phospholipase A$_2$.
[a]Mean and standard deviation for quantitative variables or count and percent for categorical variables.
[b]Criteria used for matching.
[c]Model adjusted for storage duration of serum.
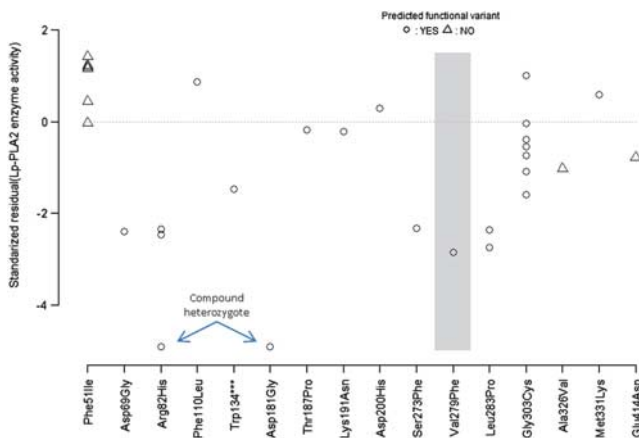


**Figure 2** Lipoprotein-associated phospholipase A$_2$ (Lp-PLA2) activity in carriers of rare non-synonymous variants. Enzyme activity is expressed as standardized residuals (variance of one) adjusting for storage duration of serum and sex. Circles correspond to carriers of variants predicted to be functional by Sorting Intolerant from Tolerant (SIFT) or Polymorphism Phenotyping (PolyPhen) and triangles otherwise. Gray background indicates the Asian null variant known from the literature.
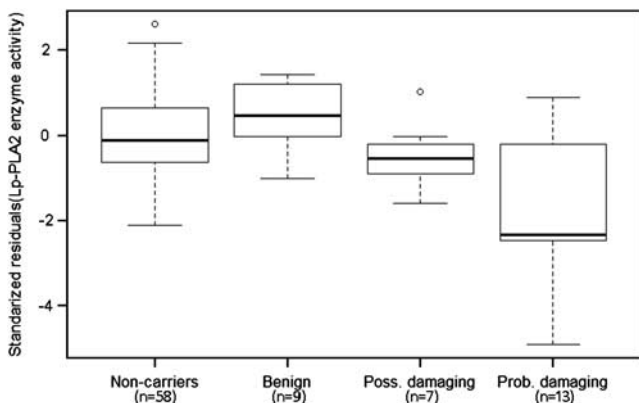


**Figure 3** Relationship between Polymorphism Phenotyping (Poly-Phen) functional prediction and lipoprotein-associated phospholipase A$_2$ (Lp-PLA2) activity adjusted for storage duration of serum and sex.

SIFT and PolyPhen were predominantly in agreement with respect to predicted impact on the protein structure and were also generally concordant with a qualitative structural assessment utilizing the crystal structure of Lp-PLA2. The one prediction that differed between the automated and the qualitative structural approach was that of the Lys191Asn. Both SIFT and PolyPhen predicted this to be functional, but the structural assessment did not. Lp-PLA2 activity levels showed that this variant had no impact on enzyme activity, though an impact on enzyme–lipoprotein interactions cannot be ruled out. Overall, in this instance, the structural predictions worked relatively well in predicting the impact on enzyme activity.

None of the common variants were associated with the cardiovascular traits tested. These results are in agreement with a recent meta-analysis by Casas *et al.*,[4] where they evaluated these common coding SNPs in 26 118 subjects and found only Ala379Val to be associated with a very modest increase in Lp-PLA2 activity (VV 7.2% higher than AA) and no association with cardiovascular-related traits, angiographic coronary disease or with coronary heart disease events. Aggregate analysis of the rare nsSNPs also failed to identify any statistically significant associations with any of the cardiovascular-related traits examined. Psychiatric traits were also not associated in either the common variant or rare aggregated analysis. However, it is difficult to conclude that no such associations exist, as these analyses have low power given the limited number of carriers and the modest relationships between Lp-PLA2 activity and such cardiovascular traits found in other studies.[1] Anecdotally, the single compound heterozygous individual observed in this study had an unremarkable clinical background, though was only 39 years old at time of entry into the study. These data would suggest that natural complete or partial deficiency in Lp-PLA2 is not clearly detrimental to human health. However, a much large study would be required to confirm this.

We attempted to genotype the nsSNPs identified via sequencing in the remaining 4188 CoLaus samples; however, most of the rare variants could not be accurately

genotyped (determined based on the comparison with the sequence-based genotypes in the overlapping portion). Given the limitations of current genotype calling algorithms for rare variants and the preponderance of singletons (most of which would not likely be observed again within another 4000 unrelated subjects), it was clear that sequencing would be required to investigate this further. This potentially has broad implications for following up deep re-sequencing studies. Many of the most functionally relevant variants will not be amenable to array-based or other standard forms of large-scale genotyping. They will also fall below the frequencies that would allow them to be imputed. If the frequency range of functional variants in *PLA2G7* proves to be the norm, then extensive high depth sequencing will be required for the investigation of rare variants in other genes.

We observed that there is relatively limited value in investigating the impact of Lp-PLA2-altering variations on disease risk or related traits in populations of European ancestry. Instead, it would be much more productive to focus on studies in East Asian populations where the functionally validated null Val279Phe variant is much more common than the cumulative rare variants in Europeans. We note the similarity to *PCSK9* and the demonstration of its influence on low-density lipoprotein cholesterol and cardiovascular disease via the functional Y142X and C679X variants present at a combined frequency of 2% in African Americans compared with <0.1% in Europeans.[24–26] Current population-based and case collections are mainly of European ancestry. Such findings support the development of similar resources in a wide range of populations, as is being attempted by the 1000 Genomes Project.[27] Finally, given the very low frequency of functional variants in *PLA2G7*, compound heterozygotes or homozygotes in the phase III trials for darapladib will be extremely rare in Caucasian populations, ∼1 in 125 000. These individuals who will be unlikely to benefit from darapladib can be detected based on very low enzyme activities, and verified by sequencing.

## Conflict of interest

Authors are all full time employees of GlaxoSmithkline. GlaxoSmithKline has an Lp-PLA$_2$ Inhibitor currently in development.

## Note added in proof

After this work was accepted, a related paper that describes the protective effect of the PLA2G7 Asian null variant V279F on coronary artery disease in Korean males was accepted for publication (Jang *et al*, *Plos ONE* 2011; **6**: 1–7).

## References

1 Thompson A, Gao P, Orfei L, Watson S, Di AE, Kaptoge S *et al.* Lipoprotein-associated phospholipase A(2) and risk of coronary disease, stroke, and mortality: collaborative analysis of 32 prospective studies. *Lancet* 2010; **375**: 1536–1544.

2 Serruys PW, Garcia-Garcia HM, Buszman P, Erne P, Verheye S, Aschermann M *et al.* Effects of the direct lipoprotein-associated phospholipase A(2) inhibitor darapladib on human coronary atherosclerotic plaque. *Circulation* 2008; **118**: 1172–1182.

3 Jang Y, Kim OY, Koh SJ, Chae JS, Ko YG, Kim JY *et al.* The Val279Phe variant of the lipoprotein-associated phospholipase A2 gene is associated with catalytic activities and cardiovascular disease in Korean men. *J Clin Endocrinol Metab* 2006; **91**: 3521–3527.

4 Casas JP, Ninio E, Panayiotou A, Palmen J, Cooper JA, Ricketts SL *et al.* PLA2G7 genotype, lipoprotein-associated phospholipase A2 activity, and coronary heart disease risk in 10 494 cases and 15 624 controls of European Ancestry. *Circulation* 2010; **121**: 2284–2293.

5 Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 2007; **39**: 513–516.

6 Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB *et al.* Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008; **40**: 592–599.

7 Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* 2008; **8**: 6.

8 Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 2006; **7**: 61–80.

9 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812–3814.

10 Sunyaev S, Ramensky V, Koch I, Lathe III W, Kondrashov AS, Bork P. Prediction of deleterious human alleles. *Hum Mol Genet* 2001; **10**: 591–597.

11 Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30**: 3894–3900.

12 Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C *et al.* The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry* 2009; **9**: 9.

13 Stephens M, Sloan JS, Robertson PD, Scheet P, Nickerson DA. Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nat Genet* 2006; **38**: 375–381.

14 Bhangale TR, Stephens M, Nickerson DA. Automating resequencing-based detection of insertion-deletion polymorphisms. *Nat Genet* 2006; **38**: 1457–1462.

15 Li B, Leal SM. Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genet* 2009; **5**: e1000481.

16 Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; **40**: 768–775.

17 Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH *et al.* LDL-cholesterol concentrations: a genome-wide association study. *Lancet* 2008; **371**: 483–491.

18 Patterson N, Price AL, Reich D. Population structure and eigen analysis. *PLoS Genet* 2006; **2**: e190.

19 Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*, 4th ed Benjamin Cummings, Longmans Green: Harlow, UK. 1996.

20 Samanta U, Bahnson BJ. Crystal structure of human plasma platelet-activating factor acetylhydrolase: structural implication to lipoprotein binding and catalysis. *J Biol Chem* 2008; **283**: 31617–31624.

21 Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A *et al.* Genes mirror geography within Europe. *Nature* 2008; **456**: 98–101.

22 Hou L, Chen S, Yu H, Lu X, Chen J, Wang L *et al.* Associations of PLA2G7 gene polymorphisms with plasma lipoprotein-associated phospholipase A2 activity and coronary heart disease in a Chinese Han population: the Beijing atherosclerosis study. *Hum Genet* 2009; **125**: 11–20.

23  Ninio E, Tregouet D, Carrier JL, Stengel D, Bickel C, Perret C *et al.* Platelet-activating factor-acetylhydrolase and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. *Hum Mol Genet* 2004; **13**: 1341–1351.

24  Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 2005; **37**: 161–165.

25  Cohen JC, Boerwinkle E, Mosley Jr TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006; **354**: 1264–1272.

26  Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin NV, Horton JD *et al.* Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. *Am J Hum Genet* 2006; **79**: 514–523.

27  Wood L, Powell D. 1000 genomes project releases data from pilot projects on path to providing database for 2500 human genomes. 2010 Ref Type: Internet Communication (http://www.1000genomes.org/docs/100621.1000GenomesPressRelease.pdf).

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)