Published in final edited form as:

DNA Repair (Amst). 2009 May 1; 8(5): 579–584. doi:10.1016/j.dnarep.2008.12.005.

Population-specific variation in haplotype composition and heterozygosity at the POLB locus

Jennifer Yamtich^{b,a}, William C. Speed^b, Eva Straka^b, Judith R. Kidd^b, Joann B. Sweasy^{b,a}, and Kenneth K. Kidd^{b,*}

^aDepartments of Therapeutic Radiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, United States

^bDepartment of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, United States

Abstract

DNA polymerase beta plays a central role in base excision repair (BER), which removes large numbers of endogenous DNA lesions from each cell on a daily basis. Little is currently known about germline polymorphisms within the POLB locus, making it difficult to study the association of variants at this locus with human diseases such as cancer. Yet, approximately thirty percent of human tumor types show variants of DNA polymerase beta. We have assessed the global frequency distributions of coding and common non-coding SNPs in and flanking the POLB gene for a total of 14 sites typed in approximately 2400 individuals from anthropologically defined human populations worldwide. We have found a marked difference between haplotype frequencies in African populations and in non-African populations.

Keywords

POLB; Alleles; Haplotypes; Single nucleotide polymorphisms; Variation; Population genetics

1. Introduction

Cancer cells have a higher mutational load than can be accounted for by the normal cellular mutation rate, leading to the long-held hypothesis that there is a mutator phenotype in cancer (reviewed in [1]). Mutants with reduced fidelity of DNA repair and replication enzymes are important for the acquisition of this phenotype (reviewed in [2]). Genomic stability is challenged by the constant presence of endogenous and exogenous DNA damaging agents, some of which result in the production of 20,000 spontaneous apurinic/aprimidinic (AP) sites per cell per day [3]. These cytotoxic lesions are the substrate for the base excision repair (BER) pathway, the DNA repair pathway primarily responsible for repairing lesions caused by reactive oxygen species and alkylating agents. DNA polymerase beta (Pol β) is a key enzyme in BER, acting as the main gap-filling enzyme and providing essential deoxyribose phosphate (dRP) lyase activity (reviewed in [4]).

Several tumor-specific single amino acid substitution mutants have been identified in pol β [5]. While it has been established that tumor-specific mutants may play a role in the etiology of cancer, little is currently known about the germline polymorphisms in pol β . dbSNP lists

189 single nucleotide polymorphisms (SNPs) in approximately 33 kb of the POLB locus, only 4 of which occur in the coding sequence (geneID:5423). Additionally, Mohrenweiser et al. [6] resequenced DNA repair genes in healthy individuals and identified three non-synonymous amino acid substitutions in pol β . To date, it is not known if any of these germ line SNPs overlap with the tumor-specific mutations. Only a few studies have tried to associate this germ-line variation in the POLB gene with altered cancer risk or survival, often with inconclusive results [7–11].

Using SNPs from both intronic and coding regions, previous studies have looked for association between variation at the POLB locus and cancer predisposition or survival. Several studies have found an association between POLB alleles and colorectal, breast, and bladder cancer risk [7–9] and lung and pancreatic cancer survival [10,11]. However, these studies are plagued by recurring limitations. First, the statistical power of the study is often limited by low frequencies of the variant alleles [7] and/or a lack of homozygotes in the study population [11], since one expects the effect of a variant to be greater in the homozygotes lacking a ‘normal’ allele. Second, many studies do not use haplotypes in their analysis, making it difficult to determine if the causative variant is the polymorphism being studied or another polymorphism in linkage disequilibrium with the genotyped marker. However, interactions of SNPs in the absence of strong LD could be missed since neither SNP alone would show as strong a signal as the haplotype. Thus haplotypes would be especially useful in studies such as in [8,9] which suggest that increased risk occurs with the presence of multiple gene variants or as in [10] which suggests that being homozygous for the variant allele of either polymorphism alters survival.

One way to resolve the above limitations is through the determination of global population-specific allele frequencies for multiple DNA repair polymorphisms and the haplotypes comprised of the allelic combinations of the polymorphisms of each gene. Such data would be very informative. First, while variants in a single gene may be important, combinations of polymorphisms in different DNA repair genes could have unique phenotypes relevant to human disease, such as cancer predisposition. Global population-specific allele and haplotype frequencies will enable the determination of which allelic combinations, both in a specific gene and in multiple DNA repair genes, are most likely to occur. Second, this information would be useful in the design of biochemistry experiments aimed at testing the function of certain polymorphisms. By determining the odds of whether any two polymorphisms for the same gene or in the same pathway are likely to occur together, one can prioritize which combinations of variants to study. Finally, haplotype analysis will provide an evolutionary perspective on how variants have arisen in a gene. The ability to root the haplotype phylogeny, using the primate data, and the ability to infer a haplotype’s relative age, based on distance from the root and global distribution, allows us to determine the likelihood of additional accumulated variation on a specific haplotype, given a constant mutation rate and absence of selective pressures. A more recent chromosome, e.g. a high-frequency recombinant haplotype seen only in North America, has had less time to accumulate additional variation than a chromosome close to the ancestral haplotype, found at moderate worldwide frequency.

As an initial study toward a full understanding of POLB variation globally, we have assessed the global frequency distributions of coding and common non-coding SNPs in and flanking the POLB gene for a total of 14 sites typed in approximately 2400 individuals from anthropologically defined human populations worldwide. These SNPs showed strong linkage disequilibrium. African populations had 3–4 common (>10%) haplotypes, while all non-African populations had a single common (generally >80%) haplotype and some populations had 1–2 others at >5%. The common non-African haplotype occurred at 20% or less in African populations. Three previously identified coding SNPs were not seen in any individuals worldwide, while two were generally low frequency and occurred in specific populations.

Further resequencing of DNA repair genes in African individuals is warranted, as our results showed that several haplotypes common in African populations are rare or unobserved outside of Africa.

2. Methods

2.1. DNA sample collection and purification

The 2442 samples studied are listed in Table 1 by population name and grouped by geographic region. All samples were collected from apparently healthy adults; except in rare cases, no phenotype data were collected. These anonymous samples were collected with informed consent under protocols approved by the relevant committees at the various institutions and countries involved. DNA was extracted using standard phenol/chloroform purification from lymphoblastoid cell lines established and/or maintained by JRK at Yale.

2.2. SNP selection and analysis

SNPs were chosen based on several criteria: their worldwide frequency, their frequency distribution among populations, and their putative functional relevance (coding SNPs, 3' UTR SNPs, etc.). All SNPs were genotyped by the TaqMan method (Applied Biosystems, www.appliedbiosystems.com) using commercial reagents and protocols. Four of the SNPs (Gln8Arg, Lys289Met, rs2953993, and rs3136804) required assays to be custom designed by Applied Biosystems (primer/probe sequences available on request). Ancestral state was determined by genotyping of 15 primates (3 gorilla, 3 chimpanzee, 3 orangutan, 3 bonobo, 3 gibbons) and comparison of human sequence to chimpanzee genomic reference sequence (genome.ucsc.edu).

2.3. PCR-RFLP analysis

Restriction fragment length polymorphism PCR (PCR-RFLP) was used to confirm the genotypes of the Lys289Met polymorphism of 89 samples of European American ancestry as in [9]. Briefly, each 20 μ l PCR reaction contained 10 ng genomic DNA, 1.25 U Taq polymerase in 1 $\times$ PCR buffer, 1.5 mM MgCl₂, 50 μ M dNTPs and 250 nM each primer. Forward primer sequence was 5' -GTCATTTCTGCCTGGGCTAAG-3' and reverse primer sequence was 5' -ATGCATTAAATCCTCTCTCGT-3'. Thermal cycling conditions began with an initial denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 30 s, with a final extension at 72 °C for 5 min. Restriction digest was performed with 5 U of *Eco*NI restriction enzyme on 10 μ l of the 568-bp PCR product for one hour at 37 °C. Digestion products were analyzed on a 2% agarose gel.

2.4. Statistical analysis

All SNPs were analyzed by population for significant deviations from Hardy–Weinberg equilibrium using a simple chi-square statistic, by locus per population. Those systems failing this test at a nominal $p < 0.01$, generally due to low frequency of one allele, are then tested using a Monte Carlo permutation test with the program HWSim (<http://krunch.med.yale.edu/hwsim/hwsim.txt>). Individuals null for more than two genotypes were omitted from haplotype analyses to minimize errors in haplotype inference. On average fewer than 2% of individuals were omitted. Genotype data were haplotyped using PHASE 2.1 [12] using default values for 5 independent runs. Data were compared with runs done on the same dataset using fast PHASE (200 sampled haplotypes, using population information); no significant discrepancies were seen.

3. Results

We genotyped 14 SNPs in approximately 2400 individuals from 46 anthropologically defined global populations. Five of the SNPs occurred in the coding sequence of the POLB gene, six were located in intronic POLB sequence, and three were located in the genomic sequence of adjacent genes (Table 2). Twelve of the SNPs were listed in the dbSNP database, and the unlisted SNPs coding for missense mutations in POLB were identified in previous publications [6,9]. For all SNPs the typings for all non-human primates were consistent with the nucleotide seen in chimpanzees being the ancestral human allele.

Population-specific allele frequencies were calculated by gene counting for each polymorphism, assuming no null alleles (Fig. 1). After correcting for multiple tests, none of the Hardy–Weinberg tests was significant. Of the genotyped amino acid substitutions previously reported, three did not vary in our populations, indicating that they are probably not polymorphic: Gln8Arg, Tyr296Tyr (rs3136804), and Lys289Met (data not shown). PCR-RFLP confirmed the results of the probe for the Lys289Met polymorphism. The derived alleles of the remaining coding polymorphisms, Arg137Gln and Pro242Arg, were generally found at low frequency worldwide. Most populations were fixed for the 137Arg allele (Fig. 1), while the 137Gln allele was found in selected Asian and North American populations, with a maximum allele frequency of 6%. Similarly, most populations were invariant for the ancestral 242Pro allele. The 242Arg allele (Fig. 1) was found at a maximum frequency of 2.4% in European populations and in the Keralites, with a trace presence in populations with low European admixture, such as African Americans and populations of the Americas.

Most of the polymorphisms were in strong linkage disequilibrium, evidenced by the low number of common haplotypes and their evolutionary relationships, as discussed below. One, rs11996105, was not. The remaining 10 polymorphisms that were varying in our populations and in strong linkage disequilibrium were used to generate global haplotype frequencies (Fig. 2). Three to four common haplotypes with frequencies greater than 10% were identified in African populations (Fig. 2, haplotype numbers one, two, six and seven). Only one common haplotype was found in non-African populations, with frequencies ranging from 72 to 100% in non-African populations and from 1 to 40% in Africans (Fig. 2, haplotype number two). The less common alleles of the Arg137Gln (rs12678588) and Pro242Arg (rs3136797) polymorphisms were not found together and were each found on only one haplotype (Fig. 2, haplotype numbers eight and four, respectively).

Population-specific haplotype heterozygosities were also calculated for the POLB locus. Africans have an average haplotype heterozygosity of 36.1% at POLB (Fig. 3). Haplotype heterozygosity is much less out of Africa, with Europe having an average haplotype heterozygosity of less than 12% and North and South America having an average heterozygosity of 3.9% (Fig. 3).

An evolutionary tree of haplotypes was constructed using maximum parsimony, with the ancestral haplotype being defined according to the chimp allele at that location (Fig. 4). There are two highly divergent branches, each with several smaller offshoots. The 242Arg and 137Gln alleles are located on separate, highly evolutionarily divergent branches. The tree shows the most upstream SNP, rs6474388, “mutating” twice. We interpret the second occurrence as a recombination event since it [1] has the largest distance between adjacent SNPs, and [2] lies outside the isochores containing the gene (45% GC-content), with a region of low GC (30%) and high GC (60%) intervening. While which is the mutation and which is the recombinant is strictly ambiguous from these data, the population distribution argues that the mutation occurred in the ancestry of haplotype number six, which is the most frequent

haplotype in Africa and haplotype number three arose as a crossover outside of Africa, where it is seen sporadically at low frequencies.

4. Discussion

We genotyped 14 polymorphisms in approximately 2400 individuals from 46 global populations and used this to calculate global allele and haplotype frequencies. Haplotype frequencies at the POLB locus showed the expected higher heterozygosities in African populations, with lower in non-African populations. However, the high frequency of only one haplotype and concomitant low heterozygosity in all non-African populations is atypical when compared to other loci studied on these populations [14]. The derived alleles of the coding polymorphisms Arg137Gln (rs12678588) and Pro242Arg (rs3136797) are found on only one haplotype apiece, and these are evolutionarily divergent. Our results show that resequencing of the POLB gene in Africans is warranted, as previous resequencing efforts could have missed polymorphisms important for human health located on haplotypes that are common in Africa but rare out of Africa.

Our results show that three previously identified “polymorphisms” are invariant in over 2400 individuals of our genotyped populations. Two of these SNPs were originally identified in the NIH Polymorphism Discovery Resource (PDR90) [13] at low frequency: the Arg allele of the Gln8Arg missense substitution was identified at a frequency of 0.01 [6] and the less common allele of the Tyr296Tyr synonymous substitution was originally identified by the Environmental Genome Project [13] at an allele frequency of 0.011. It is likely that these variants represent extremely rare, possibly individual, variants, or were identified as sequencing glitches in individual samples. Failure to see these alleles in our samples is not surprising as it confirms their status as globally rare alleles, not strictly polymorphisms (frequency >0.01). However, this is not the case for the Lys289Met variant. In this case, the 289Met allele was reported by Sliwinski and colleagues [9] to occur at high frequencies, both in their cancer patients and in their cancer-free controls (allele frequencies of 0.48 and 0.35, respectively). This contradicted previous reports, which identified the 289Met allele as a tumor-specific mutation found in colon adenocarcinoma [15]. Additionally, previous work in our lab has shown that pol β with the 289Met amino acid change is a sequence-specific mutator [16] whose expression in cell culture results in cellular transformation [17]. Our inability to identify even a single heterozygous individual in the approximately 2400 unrelated individuals genotyped, both using the TaqMan assay and PCR-RFLP analysis as confirmation, supports the initial finding that Lys289Met is a tumor-specific mutation and not a polymorphism.

The results from this study have several implications regarding biochemical analysis of polymorphisms in pol β function. First, previously identified polymorphisms (Gln8Arg, Tyr296Tyr) were not observed, suggesting that they may not be frequent enough to have a significant impact on human health. Second, the 242Arg and 137Gln alleles at rs3136797 and rs12678588, respectively, are located on separate haplotypes, making it unlikely for both derived alleles to occur in the same protein or person at appreciable frequencies. In this case, biochemical studies looking at the polymorphisms singly would be appropriate. Finally, knowledge of haplotype evolution provides insight into the chance that a secondary mutation exists. Due to the evolutionary divergence between the haplotypes containing the 242Arg and 137Gln alleles, it is possible that additional differences have accumulated between these two divergent evolutionary branches of the haplotype tree.

Knowledge of global haplotype frequencies and evolution is also applicable to the interpretation of association studies. For example, in the paper by Li et al. [10], the authors concluded that variants of intronic POLB SNPs (rs2272615, rs2953993) are associated with increased survival in patients with pancreatic cancer. However, by combining their results with

our data, two differences in interpretation become apparent. First, what they called the “variant” alleles are actually the ancestral alleles. Therefore, a revised conclusion would be that being heterozygous or homozygous for the derived alleles have moderately significant ($p = 0.019$ for rs2953993 and $p = 0.027$ for rs2272615) detrimental effects on overall survival in pancreatic cancer. Second, the haplotypes containing the derived allele with the most significant association are highly evolutionarily divergent from the ancestral haplotype (Fig. 4, haplotypes six, seven and eight), making it likely that this intronic SNP is acting as a marker for an as yet unidentified causative variant.

When resequencing genomes to identify variation, it is important to resequence individuals from populations with the highest heterozygosity and to include the most derived haplotypes. In the case of the POLB locus, the greatest heterozygosity is found in African populations. Our data show that several haplotypes common in Africa are rare in non-Africans (Fig. 2, haplotype numbers one, six and seven). These haplotypes could harbor variants with clinical relevance that were missed by previous resequencing efforts focused on European or admixed populations. Resequencing in Africans would also identify any additional non-synonymous differences between evolutionarily divergent haplotypes, such as those containing the 137Gln and 242Arg alleles. Identification of these variants would aid in the interpretation of genotypic association studies, such as the previously mentioned work on pancreatic cancer and base excision repair (BER) polymorphisms by Li et al. [10]. With this increased understanding of variation at the POLB locus, it will then be essential to conduct a similar analysis of different BER and DNA repair genes in order to understand the complex interplay between multiple variants in a given functional pathway. By identifying which potentially detrimental variants may occur together in individuals, we can begin to determine which interactions contribute to individual differences in cancer predisposition and response to treatment.

With these goals in mind, the use of large, well-defined global samples with limited admixture for the determination of population-specific allele and haplotype frequencies is essential. The samples used in this study are ideal. These samples represent one of the largest collections of global human genetic diversity with multiple samples from most major parts of the world. They have been extensively studied for thousands of genetic markers so that these samples are well characterized and the relationships of the populations they represent are clear [18,19]. Additionally, use of the same samples for the genotyping of multiple loci allows for the calculation of the frequency with which multiple variants occur in the same individual. This is essential in understanding how multiple variants of small effect combine to influence an individual's health.

Acknowledgments

This work was funded, in part, by National Institute of Health grants GM057672 to Kenneth K. Kidd and NCICA16038 funded to Joann B. Sweasy. None of the authors has any conflicts of interest related to the data presented here. We thank all the colleagues who helped us assemble the population samples, the Coriell Institute for Medical Research (NIGMS Human Genetic Cell Repository), and The National Laboratory for the Genetics of Israeli Populations at Tel-Aviv University. Special thanks are due the many hundreds of individuals from these populations who volunteered to give blood samples for studies such as this.

References

1. LAF L. Cancer Res 2001;61:3230–3239. [PubMed: 11309271]
2. Venkatesan RN, Bielas JH, Loeb LA. DNA Repair (Amst) 2006;5:294–302. [PubMed: 16359931]
3. Barnes DE, Lindahl T. Annu. Rev. Genet 2004;38:445–476. [PubMed: 15568983]
4. Beard WA, Wilson SH. Chem. Rev 2006;106:361–382. [PubMed: 16464010]
5. Starcevic D, Dalal S, Sweasy JB. Cell Cycle 2004;3:998–1001. [PubMed: 15280658]
6. Mohrenweiser HW, Xi T, Vazquez-Matias J, Jones IM.

7. Moreno V, Gemignani F, Landi S, Gioia-Patricola L, Chabrier A, Blanco I, Gonzalez S, Guino E, Capella G, Canzian F. Clin. Cancer Res 2006;12:2101–2108. [PubMed: 16609022]
8. Figueroa JD, Malats N, Real FX, Silverman D, Kogevinas M, Chanock S, Welch R, Dosemeci M, Tardon A, Serra C, Carrato A, Garcia-Closas R, Castano-Vinyals G, Rothman N, Garcia-Closas M. Hum. Genet 2007;121:233–242. [PubMed: 17203305]
9. Sliwinski T, Ziemba P, Morawiec Z, Kowalski M, Zadrozny M, Blasiak J. Breast Cancer Res. Treat 2007;103:161–166. [PubMed: 17131038]
10. Li D, Li Y, Jiao L, Chang DZ, Beinart G, Wolff RA, Evans DB, Hassan MM, Abbruzzese JL. Int. J. Cancer 2007;120:1748–1754. [PubMed: 17230526]
11. Matakidou A, el Galta R, Webb EL, Rudd MF, Bridle H. Hum. Mol. Genet 2007;16:2333–2340. [PubMed: 17855454]
12. Stephens M, Donnelly P. Am. J. Hum. Genet 2003;73:1162–1169. [PubMed: 14574645]
13. SNPs, N. NIEHS Environmental Genome Project. Seattle, WA: University of Washington;
14. Kidd KK, Pakstis AJ, Speed WC, Kidd JR. J. Hered 2004;95:406–420. [PubMed: 15388768]
15. Wang L, Patel U, Ghosh L, Banerjee S. Cancer Res 1992;52:4824–4827. [PubMed: 1511447]
16. Lang T, Maitra M, Starcevic D, Li SX, Sweasy JB. Proc. Natl. Acad. Sci. U.S.A 2004;101:6074–6079. [PubMed: 15075389]
17. Sweasy JB, Lang T, Starcevic D, Sun KW, Lai CC, Dimaio D, Dalal S. Proc. Natl. Acad. Sci. U.S.A 2005;102:14350–14355. [PubMed: 16179390]
18. Kidd, KK.; Kidd, JR. Human genetic variation of medical significance. In: Stearns, SCJCK., editor. Evolution in Health and Disease. Vol. 2nd ed.. New York: Oxford University Press; 2008.
19. Tishkoff SA, Kidd KK. Nat. Genet 2004;36:S21–S27. [PubMed: 15507999]

Fig. 1. Global population-specific allele frequencies of POLB SNPs. Populations are listed across the x-axis (for a complete listing of populations, see Table 1) and the corresponding allele frequency for each genotyped SNP is shown on the y-axis. The allele for which the frequency is shown is listed in the box next to the frequency graph. Symbols with red markers represent the coding polymorphisms, while symbols with blue marker represent polymorphisms used in published cancer association studies. Three previously identified polymorphisms (Gln8Arg, Lys289Met and rs3136804) were invariant in our samples and excluded from the graph.

Fig. 2. Global population-specific haplotype frequencies at the POLB locus. Populations are listed along the y-axis and haplotype frequencies are shown across the x-axis. Each haplotype is color-coded, with the size representing the relative frequency in each population. Haplotypes were constructed using PHASE 2.1 (12). Haplotypes seen at global frequencies less than 0.2% were combined in the residual category. For ease of comparison, the remaining eight more common haplotypes are numbered as in Fig. 4, with the ancestral haplotype (#1) first. The non-synonymous nucleotide changes are noted in red in haplotypes 4 and 8.

Fig. 3.
Haplotype heterozygosity at the POLB locus. Populations are listed across the x-axis and the corresponding haplotype heterozygosities are listed along the y-axis.

POLB Haplotype Evolution

Fig. 4.

Proposed haplotype evolution of the POLB locus. The proposed evolution of the eight most common POLB haplotypes is shown. The ancestral haplotype was determined as described in Section 2. Mutations new to a particular haplotype are underlined, while mutations that occur in a prior haplotype but are different from ancestral are in bold. Mutations whose origin is ambiguous have a subset question mark, indicating that one position represents a new mutation and the other most likely a crossover product. Haplotypes #4 and #8 contain the derived alleles of the two coding polymorphisms (Pro242Arg and Arg137Gln, respectively). For ease of comparison, haplotypes are numbered as in Fig. 2.

Table 1

Samples studied.

Geographic region	Population	N
African	Biaka Pygmy	68
	Mbuti Pygmies	39
	Yoruba	78
	Ibo	48
	Hausa	39
	Chagga	45
	Masai	21
	Sandawe	40
	African American	90
	Ethiopian Jews	32
SW Asian	Yemenite Jews	43
	Druze	102
	Samaritans	41
European	Ashkenazi Jews	81
	Adygei	54
	Chuvash	42
	Hungarians	89
	Russians, Archangel	33
	Russians, Vologda	48
	Finns	36
	Danes	51
	Irish	116
	European Americans	92
	Komi Zyriane	47
	Khanty	50
S Central Asian	Keralite, S. India	30
NE Asian/Siberian	Yakut	51
from Pacific	Nasioi Melanesians	23
Islands	Micronesians	37
E Asian	Laotians	119
	Cambodians	25
	Chinese, San Francisco	50
	Chinese, Taiwan	50
	Hakka	41
	Koreans	54
	Japanese	51
	Ami	40
	Atayal	42
	Cheyenne	56
N American	Pima, Arizona	51
	Pima, Mexico	53
	Maya	52
	Quechua	23
S American	Ticuna	65
	Rondonian Surui	47
	Karitiana	57
Total		2442

For each geographic region, the population name and number of individuals in the sample are given. Descriptions of the populations and of the specific collections of samples can be found in ALFRED (<http://alfred.med.yale.edu>).

Table 2

Summary of genotyped SNPs.

dbSNP rs#	Alleles	Strand	Ancestral allele	SNP type	NCBI build 36.1	Allele on freq. chart
rs6474388	A/G	Plus	G	I (IKKB)	42,306,311	G
NOdbSNP				M, Gln8Arg	42,315,322	
rs3136717	C/T	Plus	C	I	42,315,848	C
rs2979895	A/G	Minus (3')	A	I	42,319,364	A
rs2953993	C/T	Minus (3')	C	I	42,319,495	C
rs3136744	A/C	Plus	A	I	42,326,504	C
rs12678588	A/G	Plus	G	M, Arg137Gln	42,332,230	G
rs2953983	A/G	Minus (3')	A	I	42,332,413	G
rs2272615	A/G	Plus	G	I	42,338,205	A
rs3136797	C/G	Plus	C	M, Pro242Arg	42,345,962	G
NOdbSNP				M, Lys289Met	42,346,616	
rs3136804	C/T	Plus	T	S, Tyr296Tyr	42,346,638	
rs2073664	A/G	Minus (3')	G	S (DKK4)	42,350,943	G
rs11996105	A/G	Plus	G	Intergenic	42,357,845	G

The genotyped SNPs are listed in order of their genomic location according to the March 2006 UCSC genome assembly. SNPs are identified by their dbSNP rs number when available. Ancestral alleles listed are according to our probe. For SNP type, I = intronic, M = missense substitution, and S = synonymous substitution. Any coding substitutions are listed after the SNP type. If the SNP is located in a gene outside of the POLB locus, the genomic location is listed after the SNP type as well.