Clinical Chemistry 59:1
234–244 (2012)



# Very Low PSA Concentrations and Deletions of the *KLK3* Gene

Santiago Rodriguez,[1,2*†] Osama A. Al-Ghamdi,[1,2†] Kimberley Burrows,[1,2] Philip A.I. Guthrie,[1,2] J. Athene Lane,[2] Michael Davis,[2] Gemma Marsden,[3] Khalid K. Alharbi,[4] Angela Cox,[5] Freddie C. Hamdy,[3] David E. Neal,[6] Jenny L. Donovan,[2] and Ian N.M. Day[1,2]

**BACKGROUND:** Prostate-specific antigen (PSA), a widely used biomarker for prostate cancer (PCa), is encoded by a kallikrein gene (*KLK3*, kallikrein-related peptidase 3). Serum PSA concentrations vary in the population, with PCa patients generally showing higher PSA concentrations than control individuals, although a small proportion of individuals in the population display very low PSA concentrations. We hypothesized that very low PSA concentrations might reflect gene-inactivating mutations in *KLK3* that lead to abnormally reduced gene expression.

**METHODS:** We have sequenced all *KLK3* exons and the promoter and searched for gross deletions or duplications in *KLK3* in the 30 individuals with the lowest observed PSA concentrations in a sample of approximately 85 000 men from the Prostate Testing for Cancer and Treatment (ProtecT) study. The ProtecT study examines a community-based population of men from across the UK with little prior PSA testing.

**RESULTS:** We observed no stop codons or frameshift mutations, but we did find 30 single-base genetic variants, including 3 variants not described previously. These variants included missense variants that could be functionally inactivating and splicing variants. At this stage, however, we cannot confidently conclude whether these variants markedly lower PSA concentration or activity. More importantly, we identified 3 individuals with different large heterozygous deletions that encompass all *KLK3* exons. The absence of a functional copy of *KLK3* in these individuals is consistent with their reduced serum PSA concentrations.

**CONCLUSIONS:** The clinical interpretation of the PSA test for individuals with *KLK3* gene inactivation could lead to false-negative PSA findings used for screening, diagnosis, or monitoring of PCa.



Prostate-specific antigen (PSA)[7] is a serine protease produced specifically in the prostate. Although most PSA is found in semen, a small proportion is found in the bloodstream. The PSA test is used for diagnosis and disease monitoring and has been advocated for prostate cancer (PCa) screening *(1)*. Serum PSA concentrations are generally increased in PCa, although this test lacks specificity and the reference interval is quite wide *(2)*. In addition, there is variation in the distribution of PSA concentrations in the general population, which is due to differences in the volume of prostatic tissue and ethnicity. Drugs (e.g., finasteride) and genetic variation also influence PSA concentrations. Intraindividual variation in PSA, PSA measurement error, and their implications for early detection of PCa have also been described *(3)*.

PSA is encoded by *KLK3*[8] (kallikrein-related peptidase 3), a member of a family of 15 kallikrein genes located on human chromosome 19 *(4)*. Common single-nucleotide polymorphisms (SNPs) in *KLK3*

[1] MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE) and Bristol Genetic Epidemiology Laboratories (BGEL), University of Bristol, Bristol, UK; [2] School of Social and Community Medicine, University of Bristol, Bristol, UK; [3] Nuffield Department of Surgery, University of Oxford, Oxford, UK; [4] Clinical Laboratory Sciences Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; [5] Institute for Cancer Studies, University of Sheffield, Sheffield, UK; [6] Department of Oncology, University of Cambridge, Cambridge, UK.

* Address correspondence to this author at: MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE) and Bristol Genetic Epidemiology Laboratories (BGEL), School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. Fax +44-1179-3310132; e-mail santi.rodriguez@bristol.ac.uk.

† Equal first authorship.

The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Department of Health.

Received July 19, 2012; accepted October 9, 2012.
Previously published online at DOI: 10.1373/clinchem.2012.192815

[7] Nonstandard abbreviations: PSA, prostate-specific antigen; PCa, prostate cancer; SNP, single-nucleotide polymorphism; indel, insertion–deletion; ProtecT, Prostate Testing for Cancer and Treatment (study); NEQAS, National External Quality Assessment Service; ARCS, amplification ratio control system; ARE, androgen-responsive element; STEP, splicing and translational efficiency polymorphism.

[8] Human genes: *KLK3*, kallikrein-related peptidase 3; *TERT*, telomerase reverse transcriptase; *FGFR2*, fibroblast growth factor receptor 2; *TBX3*, T-box protein 3; *KLK15*, kallikrein-related peptidase 15; *KLK2*, kallikrein-related peptidase 2; *KLK1*, kallikrein-related peptidase 1; *ACPT*, acid phosphatase, testicular; *GPR32*, G protein–coupled receptor 32; *INS*, insulin; *KLK4*, kallikrein-related peptidase 4; *KLK9*, kallikrein-related peptidase 9; *KLK10*, kallikrein-related peptidase 10; *KLK14*, kallikrein-related peptidase 14.

have been associated with variation in PSA concentrations *(5)*, several clinical and histomorphologic features of PCa *(6)*, and susceptibility to PCa *(7)*. A genomewide association study provided early evidence of the association with PCa *(8)*. The association with PCa is controversial, however, because susceptibility to PCa could be related to PSA-based case ascertainment, with the reported risk of PCa being mediated through the effect of *KLK3* genetic variation on PSA concentrations *(9, 10)*. A subsequent genomewide association study and follow-up analysis have shown a tendency for a *KLK3* SNP allele associated with PCa risk also to be associated with high PSA concentrations *(10)*. In addition, a recent 2-stage genomewide association study showed a direct association between a novel *KLK3* SNP and PCa *(11)*. These results are consistent with a direct influence of common genetic variation in *KLK3* on PCa, because the majority of the control samples were not selected on the basis of PSA concentrations, making it less likely that the association with PCa is a reflection of an association with PSA concentrations.

The combined effect of genetic variants for 4 genes associated with PSA concentrations [*TERT* (telomerase reverse transcriptase), *FGFR2* (fibroblast growth factor receptor 2), *TBX3* (T-box protein 3), and *KLK3*] has been studied in 2 populations (in Iceland and the UK), on the assumption of a multiplicative model. For the lowest 5% of the genetic PSA distribution, the PSA concentrations were estimated to be lower than the population mean by 30%–56% in the Icelandic population and by 53%–80% in the UK population *(10)*. This combined effect, however, does not explain the very low PSA concentrations ($\leq$0.1 $\mu$g/L) found in some individuals.

Two recent studies have resequenced *KLK3* and neighboring genes from samples obtained from PCa patients and control individuals *(12, 13)*. Parikh et al. *(12)* identified 555 polymorphic loci within the region containing *KLK3* and the neighboring genes *KLK15* and *KLK2*, including 116 novel SNPs, 182 novel insertion–deletion (indel) polymorphisms, and 257 previously described loci. Klein et al. *(13)* identified 140 polymorphisms in all 15 kallikrein genes, including 38 novel SNPs. None of these studies described any major rearrangements (deletion or insertion) in *KLK3*. In addition, no study published to date has investigated genetic variation, either at the sequence level or at the level of major genetic rearrangements, in a selected group of individuals with very low PSA concentrations.

Our hypothesis is that very low serum PSA concentrations may reflect inactivating gene mutations or gross genetic rearrangements involving *KLK3*. To test this hypothesis, we conducted a detailed genetic analysis of 30 individuals with the lowest observed PSA concentrations ($\leq$0.1 $\mu$g/L) in a cohort of approximately 85 000 men from the Prostate Testing for Cancer and Treatment (ProtecT) study *(14)*.

## Materials and Methods

### INFORMATION ABOUT THE ProtecT STUDY

In the ProtecT study, approximately 85 000 men 50–69 years of age were randomly selected from 300 randomly selected general practices in 9 UK cities (broadly representing the UK white population). These men then underwent testing in PSA-testing clinics. From the results, 2000 PCa cases and 2000 controls were randomly selected to form one of the largest PCa case control studies in the world *(15, 16)*.

Men from specific primary-care centers in the cities were invited to attend a 30-min clinic appointment for a prostate check, at which time they were informed about the study and asked to consent to a PSA test. Men with an increased PSA concentration (initially $>$3.0 $\mu$g/L if 50–59 years of age and $>$4.0 $\mu$g/L if 60–69 years of age; however, the cutoff was changed to $>$3.0 $\mu$g/L for all men after 1 year) were invited for biopsy. Histologically confirmed PCa cases were identified through a combination of PSA testing, digital rectal examination, and, for men with abnormal PSA concentrations or digital rectal examination findings, a 10-core transrectal ultrasound-guided biopsy.

The inclusion criteria for this study were absence of PCa, PSA concentrations $\leq$0.1 $\mu$g/L, and availability of DNA. These criteria defined a sample of 30 individuals in the ProtecT study.

### SERUM PSA MEASUREMENTS

In the ProtecT study, the testing of serum PSA concentrations was overseen by the UK National External Quality Assessment Service (NEQAS), which provided quality assessment of the PSA tests performed in all immunology departments for the different ProtecT locations. We used 2 different immunoassays [ADVIA Centaur (Siemens Healthcare Diagnostics) and Elecsys (Roche)] for repeated PSA measurements in all 30 selected individuals. PSA measurements with the ADVIA Centaur and Elecsys systems were performed at Southmead Hospital, Bristol, UK, and in accordance with the recommendations of the 2 immunoassays' manufacturers. Both methods meet the criteria for acceptable performance, as determined by the Centre for Evidence-based Purchasing (National Health Service Purchasing and Supply Agency), with respect to accuracy and equimolarity (equivalent abilities to detect the free and complexed forms of PSA) *(17)*. The degree of agreement between the 2 methods was assessed with Pearson correlation coefficients.

#### PCR AMPLIFICATION AND SEQUENCING

We sequenced all exons and the regulatory region of *KLK3* to search for genetic defects in *KLK3* that would lead to abnormal PSA concentrations. We designed 4 long-PCR assays encompassing *KLK3* and its regulatory regions (for primers and PCR conditions, see Appendix 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue1). Long PCRs were followed by nested PCRs to generate PCR products, which were then outsourced for sequencing to GATC Biotech (http://www.gatc-biotech.com). The quality of chromatograms was analyzed before the alignment of all sequences to a reference sequence (see Appendix 2 in the online Data Supplement).

#### ASSESSMENT OF THE EFFECTS OF THE OBSERVED VARIANTS

Details of the methods used to assess the effects of the observed variants are described in Appendix 3 in the online Data Supplement. NNSplice, SpliceView, SplicePort, and Human Splicing Finder were used for predicting splice sites, and the SNP-prediction tools SIFT, PolyPhen-2, SNPs3D, PMut, and SNPs&GO were used to assess for nonsynonymous variants).

#### GROSS REARRANGEMENTS

We also looked for gross rearrangements (deletions and duplications) in the *KLK3* gene by means of a high-throughput ratiometric method, the amplification ratio control system (ARCS), which we developed and validated in our laboratory to genotype copy number variants *(18)*. This method is based on a novel PCR protocol that analyzes the ratio of the number of copies of a gene with a variable copy number to a single-copy reference gene. The amplicons for the target and reference genes were designed to have melting temperatures that differed by a few degrees Celsius to enable their differentiation in a melting assay. The ratio of the copy number of the target gene to the reference gene is inferred from the change in fluorescence contributed by each gene as it undergoes melting. We designed ARCS assays for *KLK3* exons and neighboring genes both upstream and downstream (see Appendix 4 in the online Data Supplement for primers and PCR conditions).

## Results

#### SERUM PSA CONCENTRATIONS

Table 1 shows serum PSA concentrations for all 30 ProtecT participants measured with the Centaur and Elecsys immunoassays for the same samples, compared with the initial PSA concentrations. The measurements obtained with the 2 immunoassays were in close agreement (Pearson *r*, 0.987). All measurements were <0.5 μg/L, with the exception of 2 individuals (S5 and S17), whose reassay results were approximately 0.8 μg/L and 1.2 μg/L, respectively. The data for these 2 individuals were also included in our analysis. Note that the PSA concentrations in both instances are much lower (approximately 1 to 1.5 SDs lower) than the population mean observed for healthy men of different ages [mean (SD), 2.9 (1.7) μg/L for men 60–69 years of age and 2.0 (1.2) μg/L for men 50–59 years of age] *(19)*.

**Table 1. Comparison of the PSA concentrations observed for each of the 30 samples with 3 independent measurements.[a]**

| | PSA concentration, μg/L | | |
|---|---|---|---|
| Participant | Original | Centaur | Elecsys |
| S1 | 0.08 | 0.13 | 0.12 |
| S2 | 0.003 | 0.49 | 0.51 |
| S3 | 0.09 | 0.12 | 0.12 |
| S4 | 0.07 | 0.26 | 0.34 |
| S5 | 0.06 | 0.69 | 0.85 |
| S6 | 0.05 | 0.31 | 0.41 |
| S7 | 0.08 | 0.09 | 0.11 |
| S8 | 0.09 | 0.18 | 0.13 |
| S9 | 0.07 | 0.44 | 0.52 |
| S10 | 0.08 | 0.22 | 0.28 |
| S11 | 0.08 | 0.38 | 0.43 |
| S12 | 0.06 | 0.26 | 0.34 |
| S13 | 0.05 | 0.22 | 0.23 |
| S14 | 0.09 | 0.09 | 0.09 |
| S15 | 0.10 | 0.16 | 0.08 |
| S16 | 0.10 | 0.17 | 0.15 |
| S17 | 0.10 | 1.06 | 1.24 |
| S18 | 0.10 | 0.24 | 0.27 |
| S19 | 0.10 | 0.12 | 0.22 |
| S20 | 0.10 | 0.16 | 0.23 |
| S21 | 0.10 | 0.18 | 0.20 |
| S22 | 0.10 | 0.16 | 0.20 |
| S23 | 0.10 | 0.16 | 0.17 |
| S24 | 0.10 | 0.12 | 0.13 |
| S25 | 0.10 | 0.05 | 0.05 |
| S26 | 0.10 | 0.08 | 0.09 |
| S27 | 0.10 | 0.10 | 0.11 |
| S28 | 0.10 | 0.14 | 0.18 |
| S29 | 0.10 | 0.11 | 0.14 |
| S30 | 0.10 | 0.26 | 0.36 |

[a] Original PSA values from the ProtecT study and PSA concentrations obtained via 2 different immunoassay systems: ADVIA Centaur assay (Siemens Healthcare Diagnostics) and Elecsys total PSA immunoassay (Roche).

**Table 2. *KLK3* variants found in 30 samples obtained from the ProtecT biorepository.**

| *KLK3* fragment | dbSNP ID | Position (build 37) | Alleles | Type | Coding | Genotype sequencing (for 30 samples), n | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Wild type | Heterozygous mutants | Homozygous mutants |
| ARE III | | 19:51353907 | C/A | Upstream | | 24 | 1 | None |
| | rs57268074 | 19:51353908–51353907 | _/C | Upstream, insertion | | 9 | 10 | 6 |
| | rs11084033 | 19:51353955 | C/A | Upstream | | 19 | 3 | 2 |
| ARE II | rs3760722 | 19:51357816 | C/T | Upstream | | 14 | 3 | None |
| | Unreported | 19:51357827 | T/C | Upstream | | 5 | 12 | 2 |
| ARE I | rs17526278 | 19:51357887 | G/A | Upstream | | 23 | 4 | None |
| | rs4802754 | 19:51357920 | G/A | Upstream | | 18 | 8 | 1 |
| | rs17554958 | 19:51357967–51357966 | _/A | Upstream, insertion | | 18 | 8 | 1 |
| | rs266882 | 19:51358013 | G/A | Upstream | | 7 | 15 | 5 |
| Exon 2 | rs11573 | 19:51359497 | T/C | Synonymous | Gly | 11 | 13 | None |
| | rs1135766 | 19:51359503 | A/G | Synonymous | Ala | 11 | 13 | None |
| | rs7252245 | 19:51359566 | G/A | Synonymous | Gln | 22 | 2 | None |
| | rs174776 | 19:51359852 | T/C | Intronic | | 16 | 4 | 4 |
| Exon 3 | Unreported | 19:51361307 | C/T | Nonsynonymous | Arg/Trp | 22 | 1 | None |
| | rs12946 | 19:51361315 | C/T | Synonymous | Ser | 19 | 4 | None |
| | rs61752561 | 19:51361382 | G/A | Nonsynonymous | Asp/Asn | 22 | 1 | None |
| | rs2003783 | 19:51361472 | C/A | Nonsynonymous | Leu/Ile | 21 | 2 | None |
| | rs1810020 | 19:51361644 | A/G | Intronic | | 18 | 4 | 1 |
| Exon 4 | rs17632542 | 19:51361757 | T/C | Nonsynonymous | Ile/Thr | 17 | 5 | 1 |
| | rs111901464 | 19:51361879 | G/A | Intronic | | 22 | 1 | None |
| | rs266875 | 19:51361937 | A/G | Intronic | | 8 | 13 | 1 |
| | rs34750956 | 19:51361996 | C/T | Intronic | | 20 | 3 | None |
| Exon 5 | Unreported | 19:51362715 | T/G | Intronic | | 16 | 6 | 1 |
| | rs66592214 | 19:51362716 | T/C | Intronic | | 21 | 2 | None |
| | rs35192866 | 19:51362803 | C/T | Synonymous | Thr | 21 | 2 | None |
| | rs45588133 | 19:51362955 | G/A | 3′ UTR[a] | | 21 | 2 | None |
| | rs2659122 | 19:51363026 | C/T | 3′ UTR[a] | | 8 | 10 | 5 |
| | rs1058205 | 19:51363398 | C/T | 3′ UTR | | 10 | 8 | 5 |
| | rs1058274 | 19:51363448 | A/G | 3′ UTR | | 9 | 14 | None |
| | rs6998 | 19:51363661 | G/A | 3′ UTR | | 13 | 10 | None |

[a] Classified as a 3′ untranslated region (UTR) SNP in the ENST00000360617 transcript and as an intronic SNP in the ENST00000326003 transcript (source: Ensembl).

### GENETIC VARIATION AT THE SEQUENCE LEVEL

We detected 30 genetic variants in the *KLK3* region: 25 SNPs, 2 indels, and 3 previously unreported SNPs (Table 2). Also shown are the observed genotype frequencies in our study.

We identified 9 polymorphic loci in the regulatory regions upstream of *KLK3* (Table 2). Two were indels: an A insertion (rs17554958) in the proximal promoter and a C insertion (rs57268074) in the *KLK3* enhancer. One variant was a previously unreported C-to-A variant at an SNP locus for a reported C-to-G SNP (rs73932613) adjacent to rs17554958. Two variants were SNPs within androgen-responsive elements (ARE): rs266882 in ARE I and rs11084033 in the nonconsensus ARE VI. The 4 other SNPs upstream of *KLK3* were rs3760722, rs17526278, rs4802754, and a previously unreported SNP (minor-allele frequency, 0.44).

Nine coding variants were found in *KLK3* exons—5 synonymous and 4 nonsynonymous variants (Table 2).

The nonsynonymous variants include a previously unreported C-to-T variant in exon 3, which encodes the substitution of a tryptophan residue for an arginine residue at position 77 (Arg77Trp) in PSA. Two other missense mutations were found in *KLK3* exon 3: rs61752561 (Asp102Asn), and rs2003783 (Leu132Ile). The *KLK3* exon 4 SNP (rs17632542) encodes the Ile179Thr variant in PSA. We also detected a synonymous T-to-C variant (rs11573) within *KLK3* exon 2.

The remaining 12 variants (Table 2) consisted of 7 intronic polymorphisms, 5 of which are previously reported SNPs in the 3′ untranslated region. The intronic SNPs include 2 that are previously unreported: a G-to-A variant in intron 4 and a T-to-G variant in intron 5 (see Fig. 1 in the online Data Supplement for the genotypes for all 30 variants observed in the 30 individuals).

### CONSEQUENCES FOR SPLICING

The synonymous T-to-C mutation (rs11573) is 2 bp downstream of the 3′ end of intron 1. This variant has previously been reported in the dbSNP database as a *KLK3* splice site variant, although it is not in the splice consensus sequence region (i.e., it is not within the obligatory dinucleotide essential for effective recognition of the acceptor site at the 3′ end of intron 1).

On the other hand, our results obtained with splicing-prediction tools show that 1 of the 30 sequence variants we detected (rs7252245) lies within a predicted splice site (see Table 1 in the online Data Supplement). All of the other variants we detected are located at least 10 bp away from the predicted acceptor or donor splice sites (see Table 1 in the online Data Supplement).

Table 2 in the online Data Supplement shows the FASTA sequence of *KLK3*, annotated splice sites, and SNPs identified in the present work.

### PHENOTYPIC CONSEQUENCES OF NONSYNONYMOUS VARIANTS

Table 3 shows the predicted effects of the 4 nonsynonymous variants on protein function. The novel Arg77Trp variant in exon 3 shows evidence that is has probable damaging effects. All 5 prediction tools used in this study predicted this variant as "damaging (pathological)," with the PMut tool predicting the most pathologic (i.e., maximum) effect at position 77. SNPs&GO predicted this variant to be a "disease" variant, although the reliability of this prediction was remarkably low (Reliability Index = 1). Similarly, the Asp102Asn and Leu132Ile replacements were variously described as "tolerated" (SIFT), "neutral" (PMut), and "benign" (PolyPhen-2) (Table 3). The outcomes for the SNPs&GO prediction tool for the Asp102Asn and Leu132Ile variants were in line with those for other tools. Conflicting results were observed for the

**Table 3. Nonsynonymous variants found in the present study.[a]**

| | | SNP-prediction tools | | | | |
|---|---|---|---|---|---|---|
| Variant | New/known SNP | SIFT | PolyPhen-2 | SNPs3D | PMut | SNPs&GO |
| Arg77Trp (exon 3) | New | Affects protein function | Probably damaging—Arg conserved in many other species | Probably damaging but no data given for structural effects | NN output = 0.9381 (pathologic); reliability = 8 | Disease (RI[b] = 1) |
| Asp102Asn (exon 3) | Known (rs61752561) | Tolerated | Benign—Asp in humans, Asn conserved across multiple species | Tolerated | NN output = 0.0975 (neutral); reliability = 8 | Neutral (RI = 9) |
| Leu132Ile (exon 3) | Known (rs2003783) | Tolerated | Benign—Leu in humans, Ile conserved across multiple species | Tolerated | NN output = 0.0191 (neutral); reliability = 8 | Neutral (RI = 9) |
| Ile179Thr (exon 4) | Known (rs17632542) | Affects protein function | Benign | Probably damaging—hydrophobic interaction decreased | NN output = 0.3185 (neutral); reliability = 8 | Neutral (RI = 5) |

[a] The nonsynonymous variants (1 novel and 3 previously reported variants) were input into the SNP-prediction tools SIFT, PolyPhen-2, SNPs3D, PMut, and SNPs&GO for predicting the possible effects of amino acid substitutions on protein function. The results indicated evidence for probable damaging effects of novel variant Arg77Trp, located in exon 3.
[b] RI, Reliability Index.

Ile179Thr substitution in the protein sequence. This variant was predicted as damaging by both SIFT (score, 0.01; 136 sequences queried) and SNPs3D [score, −2.45 for the 261-residue polypeptide chain (including pre- and pro-PSA) and −2.39 for the active protein comprising 238 residues] (Table 3), with SNPs3D predicting a loss of hydrophobicity at the molecular level. On the other hand, PMut ranked this variant as "neutral" (NN output, 0.3185; Reliability Index, 3), SNPs&GO predicted it as "neutral," with a moderate Reliability Index (5 of a maximum of 9), and PolyPhen-2 ranked it as "benign" (score, 0.152; sensitivity, 0.91; specificity, 0.83).

#### COMPARISON OF ALLELE FREQUENCIES

Table 4 presents the *P* values obtained for comparisons of our allele counts with European data from the dbSNP database, with data from Parikh et al. *(12)*, and with both data sets combined. Comparisons were not possible for 3 polymorphisms that no frequency data reported in either dbSNP or Parikh et al. *(12)*. These polymorphisms are the far-upstream indel rs57268074, the novel C-to-T variant in the ARE II region, and the novel C-to-T variant encoding the Arg77Trp substitution.

Note that the allele frequencies of the splice variants rs11573 and rs7252245 were not significantly different from the allele frequencies reported in dbSNP and by Parikh et al. *(12)* (see Table 4).

Only 2 variants in our resequencing data were significantly different in frequency from those in both the dbSNP database and the findings of Parikh et al. *(12)*: the intronic SNP rs174776 (T to C) and the 3′ untranslated region SNP rs1058205 (C to T). The minor allele in the latter SNP was significantly overrepresented in our sample ($P = 0.007$, compared with the dbSNP data; $P = 0.003$, compared with the data of Parikh et al. *(12)*; $P = 0.002$, compared with the 2 data sets combined). The allele frequencies of 2 additional SNPs [the intronic SNP rs1810020 (A to G) and the 3′ untranslated region SNP rs2659122 (C to T)] were significantly different from frequencies reported by Parikh et al. *(12)* but not from those reported in the dbSNP database. When data from both sets were combined, the resulting frequencies were statistically different from the allele frequencies for these 2 loci in our sample (Table 4); however, none of these significant differences withstand the Bonferroni correction for multiple testing ($\alpha_{Bonferroni} = 0.05/71 = 0.0007$).

#### PROTEIN SECONDARY STRUCTURE

Fig. 2 in the online Data Supplement shows the alignment of the nucleotide and amino acid sequences with annotated variants (our nucleotide sequence variants are represented below). The sequences are also aligned with the protein's secondary structure (obtained from the Protein Data Bank) to put the amino acid substitutions into the context of PSA structural features in which they are located. No evidence of novel genetic variation leading to changes in PSA concentrations was apparent from possible changes in PSA secondary structure.

#### LARGE DELETIONS SPANNING THE WHOLE OF *KLK3*

Three of the 30 participants with very low PSA concentrations (≤0.1 μg/L) were heterozygous for different major deletions that spanned all *KLK3* exons. Fig. 1 shows an example of the evidence for these deletions for these 3 individuals. Individual S7 showed deletion of all *KLK3* exons (see Appendix 5 in the online Data Supplement) and the neighboring gene *KLK15* (upstream of *KLK3*). There was no evidence of deletion of the downstream gene *KLK2* (exon 2). All of *KLK3* was deleted in individual S8, along with *KLK15* and *KLK1* (upstream of *KLK3*), and *KLK2* (downstream of *KLK3*) (see Appendix 5 in the online Data Supplement). Individual S23 showed evidence of deletion of all of *KLK3*, along with a long region upstream of *KLK3* that included *KLK15*, *KLK1*, *ACPT* (acid phosphatase, testicular), and *GPR32* (G protein–coupled receptor 32). The last 2 genes are outside the kallikrein gene cluster. S23 also showed partial deletion of *KLK2* (see Appendix 5 in the online Data Supplement). This finding confirmed heterozygosity for total deletion of *KLK3* in all 3 individuals. Fig. 2 shows a schematic of the extents of the 3 deletions found in this study. Each exon assay result represents independent confirmation of the occurrence of a gross deletion in the *KLK3* region in these individuals.

## Discussion

This study is the first detailed characterization of single-nucleotide genetic variations and gross rearrangements involving *KLK3* in individuals with very low PSA concentrations. Our study confirms the existence of 3 novel deletions. ARCS ratiometric analyses of *KLK3* exons and neighboring genes enabled us to determine that these deletions span the whole of *KLK3*. These deletions represent the first described human deletion mutants of *KLK3* and have important implications for interpreting PSA test results with respect to PCa for some individuals.

No frameshift mutations and no premature stop codons, which might be causes of very low PSA concentrations, were observed in our sequence data. The data on *KLK3* genetic variation in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?locusId=354) includes records of 2 previously reported indels in *KLK3* that may cause frameshift mutations. In the indel locus rs17849961, a

**Table 4. Comparison of allele frequencies observed for the resequencing data in our study and in other publications.**

| Locus | Amplicon[a] | Alleles | Allele counts, n/n | | *P* | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ProtecT[b] | dbSNP[c] | Parikh et al.[d] | (vs dbSNP)[e] | (vs Parikh et al.)[f] | (vs both)[g] |
| rs73932613 | ARE III | C/G A[h] | 49/1 | 47/3 | | 0.61 | | |
| rs57268074 | | _/C[i] | 28/22 | | | | | |
| rs11084033 | | C/A | 43/7 | 57/15 | 118/34 | 0.47 | 0.28 | 0.44 |
| rs3760722 | ARE II | C/T | 31/3 | 62/10 | 140/12 | 0.67 | 1.00 | 3.06 |
| Unreported | | C/T | 22/16 | | | | | |
| rs17526278 | ARE I | G/A | 50/4 | 62/10 | 137/11 | 0.39 | 1.00 | 0.26 |
| rs4802754 | | G/A | 44/10 | 57/15 | 105/37 | 0.92 | 0.36 | 0.46 |
| rs34823699 | | _/A[i] | 44/10 | | 100/34 | | 0.42 | |
| rs266882 | | G/A | 29/25 | 38/34 | 66/56 | 1.00 | 1.00 | 0.98 |
| rs11573 | Exon 2 | T/C | 35/13 | 42/30 | 82/60 | 0.15 | 0.09 | 0.16 |
| rs1135766 | | A/G | 35/14 | 66/50 | 85/59 | 0.12 | 0.17 | 0.20 |
| rs7252245 | | G/A | 46/2 | | 147/1 | | 0.30 | |
| rs174776 | | T/C | 26/12 | 63/9 | 134/18 | 0.03 | 0.006 | 0.007 |
| Unreported | Exon 3 | C/T | 45/1 | | | | | |
| rs12946 | | C/T | 42/4 | 70/2 | 139/13 | 0.15 | 1.00 | 0.26 |
| rs61752561 | | G/A | 45/1 | 68/4 | 148/6 | 0.67 | 0.92 | 0.65 |
| rs2003783 | | C/A | 44/2 | 63/9 | 143/11 | 0.25 | 0.74 | 0.23 |
| rs1810020 | | A/G | 40/6 | 69/3 | 144/4 | 0.16 | 0.02 | 0.02 |
| rs17632542 | Exon 4 | T/C | 39/7 | 66/6 | 131/15 | 0.39 | 0.51 | 0.49 |
| rs111901464[j] | | G/A | 45/1 | | 151/1 | | 0.95 | |
| rs266875 | | G/A | 29/15 | 118/66 | 78/72 | 0.96 | 0.14 | 0.05 |
| rs34750956 | | C/T | 43/3 | 64/8 | 136/12 | 0.61 | 0.97 | 0.65 |
| Unreported | Exon 5 | T/G | 38/8 | | 94/6 | | 0.06 | |
| rs66592214 | | T/C | 44/2 | 62/10 | 96/8 | 0.17 | 0.69 | 0.17 |
| rs35192866 | | C/T | 44/2 | 62/10 | 130/10 | 0.17 | 0.75 | 0.13 |
| rs45588133 | | G/A | 44/2 | 62/10 | 136/10 | 0.17 | 0.75 | 0.12 |
| rs2659122 | Exon 5 | C/T | 26/20 | 50/22 | 114/32 | 0.22 | 0.004 | 0.02 |
| rs1058205 | | C/T | 28/18 | 61/11 | 129/27 | 0.007 | 0.003 | 0.002 |
| rs1058274 | | A/G | 32/14 | 49/23 | 109/47 | 1.00 | 1.00 | 0.96 |
| rs6998 | | G/A | 36/10 | 48/24 | 96/58 | 0.25 | 0.07 | 0.13 |

[a] Amplicons as described in Materials and Methods.
[b] ProtecT: Resequencing data of the present study.
[c] dbSNP frequency data follow the same specifications as described in Materials and Methods and in Table 2.
[d] From supplementary data of Parikh et al. *(12)*.
[e] *P* values from $\chi^2$ comparisons with frequency data in the dbSNP database.
[f] *P* values from $\chi^2$ comparisons with frequency data of Parikh et al. *(12)*.
[g] *P* values from $\chi^2$ comparisons with frequency data from the dbSNP database and Parikh et al. *(12)* combined.
[h] Frequency data are based on the reported C/G variant; a C/A variant was detected at the same locus in our study.
[i] _/C and _/A are insertion mutations.
[j] Named as CGF_33864 in Parikh et al. *(12)*.

GC pair is replaced by a single nucleotide. The other reported frameshift mutation is rs34953540; an _/A indel polymorphism. None of these indels were found in our sample set, and no associations between these indels and serum PSA concentrations have been reported in the literature.

Alternative splicing *(20)* and intron-retention *(21)* events have been described for *KLK3*. The results

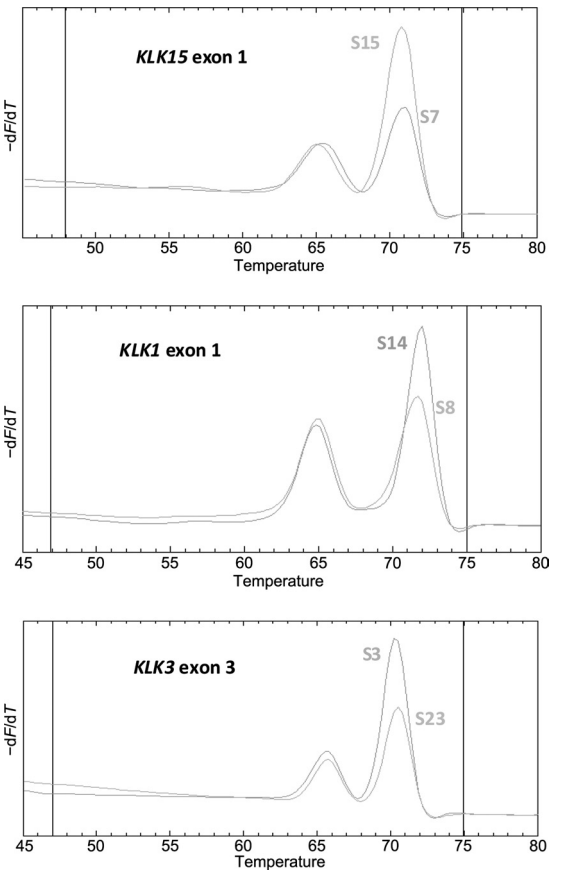


**Fig. 1. Melting profiles of 3 samples analyzed in this study.**

Each sample has 2 peaks. The peak on the left represents the AT-rich reference amplicon (typically melts at 65 °C with 200 mL/L formamide). The peak on the right represents the amplicon for *KLK3* exon 3 (melts at 70 °C with 200 mL/L formamide). The reference peaks for the 3 samples are similar, whereas the target gene (*KLK3* exon 3) shows variation. Three samples (S7, S8, and S23) show lower peaks than the control samples (S15, S14, and S3). This result is consistent with the occurrence of a heterozygous deletion in S7, S8, and S23. −d*F*/d*T*, first negative derivative of fluorescence, *F*, with respect to time, *T*.

of our analyses with splicing-prediction tools show that most of the 30 sequence variants detected in our work are not within *KLK3* acceptor or donor sites. We observed in our study 2 genetic variants within splice sites (rs11573 and rs7252245). Both are in exon 2, far apart from the putative donor and acceptor sites involved in the retention of intron 3 in the PSA rp-2 transcript *(21)*. These findings rule out the possibility of an influence of these variants on intron 3 retention. In addition, the allele frequencies of one of these SNPs (rs11573) were not significantly different from those reported in dbSNP. This finding suggests that this SNP is not a cause of very low PSA concentrations. If it were, we would expect an overrepresentation of 1 allele in our selected sample. We observed rs7252245 allele A in only 2 individuals and only in heterozygosis; therefore, it could not explain the very low PSA concentrations observed in all 30 samples.

On the other hand, the polymorphism in the polypyrimidine tract in the noncoding intron I near the 3′ splice site of the *INS* (insulin) gene has been reported to affect pre-mRNA splicing and proinsulin secretion. The A allele of the *INS* gene restriction fragment length polymorphism −23*Hph*I leads to retention of intron 1, and the extended mRNAs generate 6-fold more proinsulin in culture supernatants than natural transcripts *(22, 23)*. This new type of polymorphism has been termed "splicing and translational efficiency polymorphism" (STEP) *(24)*. A database including 3324 candidate STEPs is available online *(25)*. Therefore, it is possible that DNA sequence variation at or near splice sites within *KLK3* could explain differences in PSA concentrations; however, no STEP in the online STEP database has been identified within *KLK3 (25)*.

The finding of nonsynonymous variants raises the possibility of functional effects on PSA. We found good agreement in general among the results obtained with the SIFT, PolyPhen-2, SNPs3D, PMut, and SNPs&GO tools. An interesting finding of our work is a previously unreported missense variation (Arg77Trp in exon 3). This substitution may be functionally inactivating, but we have no evidence at this stage that it leads to very low PSA concentrations.

Therefore, we cannot confidently conclude whether any of the single-base variants detected in this work markedly lower PSA concentration or activity.

The allele frequencies observed for 2 SNPs in our study were significantly different from those reported in the dbSNP database and by Parikh et al. *(12)*; however, these significant differences may be type I errors associated with multiple testing. In addition, if these differences were real, it would be very unlikely that they accounted for the very low concentrations observed in our samples, given the relatively reduced magnitude of effect previously observed for common variants in *KLK3* and in other genes that influence serum PSA concentrations *(10)*.

Further work is needed to define the spectrum of allelic inactivation in the wider population and the precise locations of breakpoints for the deletions. We note at this stage that at least 1 carrier of the heterozygous *KLK3* mutation was fertile. Overall, the number of offspring for these 30 individuals also was consistent with the national average. The same *KLK3* deletion carrier

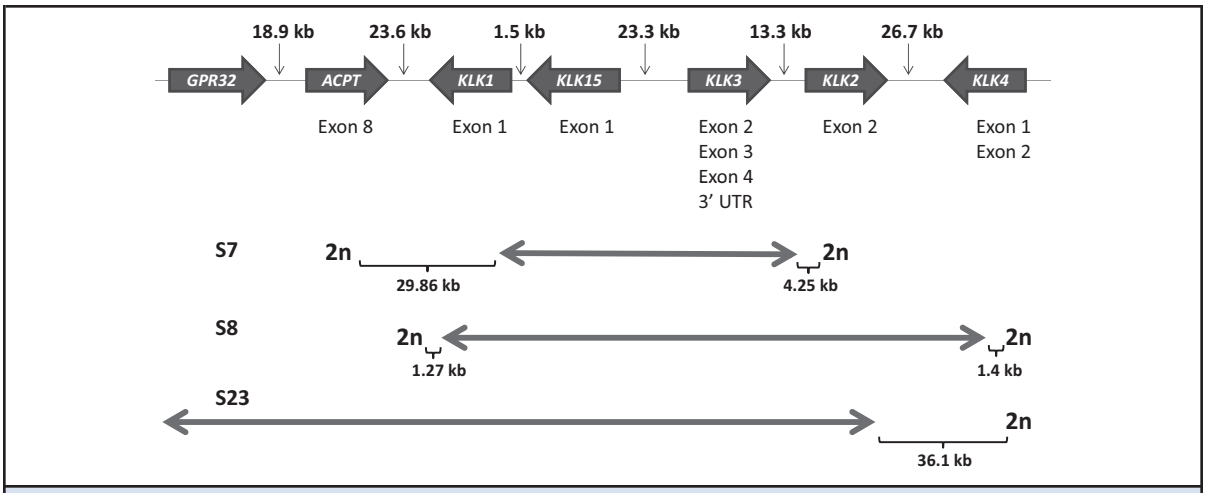


**Fig. 2. Schematic of the *KLK3* gene and neighboring kallikrein genes (*KLK15*, *KLK2*, and *KLK4*) and *ACPT*.**

The extents of the heterozygous deletions detected in samples S7, S8, and S23 are represented as red lines [2n, normal copy number (i.e., presence of 2 copies)]. Loci genotyped by ARCS are listed below each gene. The 3 deletions in these individuals all span the whole of *KLK3* and *KLK15*.

had also undergone transurethral resection for benign prostatic disease.

A point to add is that the individuals in whom we have identified gross deletions in *KLK3* display very low PSA concentrations, a finding that raises questions about the actual activity of the enzyme encoded by the nondeleted allele. Results from our study show that there is not a good correlation with a dosage effect. A correlation would be expected if heterozygotes for the deletion had PSA concentrations intermediate between zero and the concentrations in those homozygous for nondeletion. Our results did not allow us to determine whether these individuals show low expression of the other allele or whether it might also be affected. It is possible that other alterations (e.g., changes in gene regulation) are affecting the other allele in these individuals with very low PSA concentrations but that they were not detectable in our survey. One hypothesis for the very low PSA concentrations found in the carriers of a single copy of one of these deletions could be a dominant loss-of-function effect. An example is retinitis pigmentosa, in which there are 2 wild-type alleles (a high-expression allele and a low-expression allele). The combination of a mutant allele with a high-expression allele produces no disease phenotype, whereas the presence of a mutant allele and the low-expression allele produces protein concentrations below those required for normal function, thereby causing the disease *(26)*.

On the other hand, results obtained with exome chips (ftp://share.sph.umich.edu/exomeChip/ProposedContent/codingContent) have indicated the occurrence of stop codons in several *KLK* genes, including *KLK4*, *KLK9*, *KLK10*, and *KLK14*. That would be consistent with fewer functional constraints in this cluster, which would increase the frequency of gene inactivations compared with other genomic regions.

The PSA test is routinely used to diagnose and monitor PCa, with at least 40 million PSA tests having been carried out worldwide in 2007 alone *(27)*. Our findings confirmed our hypothesis that very low PSA concentrations could be caused by a major mutation in *KLK3* in some individuals. Such low expressors may be at a disadvantage in PSA-based biomarker diagnosis, monitoring, and screening tests, in that their haploinsufficiency leads to false-negative results, and therefore false reassurance. Given that SNP variation in *KLK3* seems to mark PCa risk as well as PSA concentrations *(10)*, major allelic mutations may also participate in affecting PCa risk or benign prostatic disease.

Although our study has shown a relatively low frequency of deletions and our analysis has been based on a sample of individuals with extreme PSA concentrations, it is important to recognize that deletions can occur and produce false-negative results in PCa screening programs. Future studies with large population cohorts are required to determine the frequency of *KLK3* deletions in the general population.

The identification of deletion mutants opens up the possibility of new ways to explore the role of the *KLK3* gene in PSA function. These results enable analyses of the effect of *KLK3* dose on its expression. They also enable the characterization of the phenotypic consequences associated with *KLK3* deletion. In addition, our observations reinforce the concept that copy num-

ber variants play important roles in diverse medical contexts, in this instance regarding a leading biomarker and a possible factor contributing to benign prostatic hyperplasia and PCa.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

**Authors' Disclosures or Potential Conflicts of Interest:** *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

**Employment or Leadership:** F.C. Hamdy, University of Oxford.
**Consultant or Advisory Role:** D.E. Neal, International Health Technology Ltd.
**Stock Ownership:** None declared.
**Honoraria:** None declared.
**Research Funding:** Cancer Research UK; S. Rodriguez, Cancer Research UK Small Grant (no. C30152/A11314); O.A. Al-Ghamdi, overseas PhD studentship from King Saud University; A. Cox, US Department of Defense (award no. W81XWH-04-1-0280) and Yorkshire Cancer Research; F.C. Hamdy, National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, UK Department of Health (DOH) NIHR Healthy Technology Assessment Programme (HTA) (ProtecT grant), and National Cancer Research Institute (NCRI)/Medical Research Council (MRC) (ProMPT grant); D.E. Neal, NIHR Cambridge Biomedical Research Centre, UK DOH HTA (ProtecT grant) (HTA 96/20/99), and NCRI/MRC (ProMPT grant) (G0500966/75466); J.L. Donovan, NIHR Cambridge Biomedical Research Centre, UK DOH HTA (ProtecT grant), and NCRI/MRC (ProMPT grant).
**Expert Testimony:** None declared.
**Patents:** None declared.

**Role of Sponsor:** The sponsor oversaw the conduct of the trial but played no role in the design of study, choice of enrolled patients, collection, review, and interpretation of data, or approval of manuscript.

**Acknowledgments:** We thank M. Gillett and the staff of Clinical Biochemistry laboratories, Southmead Hospital (Bristol), for conducting PSA assays. We are grateful to study volunteers for their participation and to staff at the Wellcome Trust Clinical Research Facility, Addenbrooke's Clinical Research Centre, Cambridge, for their help in conducting the study. The authors acknowledge the tremendous contributions of all members of the ProtecT study research group, including P. Bollina, D. Cooper, A. Doble, A. Doherty, E. Elliott, D. Gillatt, P. Herbert, J. Howson, R. Kockelbergh, H. Kynaston, N. Lyons, H. Moody, P. Powell, S. Prescott, and P. Thompson.

## References

1. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat Rev Cancer 2008;8:268–78.
2. Oesterling JE, Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM. Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges. JAMA 1993;270:860–4.
3. Yan Y. Intraindividual variation of prostate specific antigen measurement and implications for early detection of prostate carcinoma. Cancer 2001;92:776–80.
4. Lawrence MG, Lai J, Clements JA. Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. Endocr Rev 2010;31:407–46.
5. Cramer SD, Chang BL, Rao A, Hawkins GA, Zheng SL, Wade WN, et al. Association between genetic polymorphisms in the prostate-specific antigen gene promoter and serum prostate-specific antigen levels. J Natl Cancer Inst 2003;95:1044–53.
6. Cramer SD, Sun J, Zheng SL, Xu J, Peehl DM. Association of prostate-specific antigen promoter genotype with clinical and histopathologic features of prostate cancer. Cancer Epidemiol Biomarkers Prev 2008;17:2451–7.
7. Lai J, Kedda MA, Hinze K, Smith RL, Yaxley J, Spurdle AB, et al. PSA/KLK3 AREI promoter polymorphism alters androgen receptor binding and is associated with prostate cancer susceptibility. Carcinogenesis 2007;28:1032–9.
8. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008;40:316–21.
9. Ahn J, Berndt SI, Wacholder S, Kraft P, Kibel AS, Yeager M, et al. Variation in KLK genes, prostate-specific antigen and risk of prostate cancer. Nat Genet 2008;40:1032–4.
10. Gudmundsson J, Besenbacher S, Sulem P, Gudbjartsson DF, Olafsson I, Arinbjarnarson S, et al. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med 2010;2:62ra92.
11. Kote-Jarai Z, Amin Al Olama A, Leongamornlert DA, Tymrakiewicz M, Saunders J, Guy M, et al. Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. Hum Genet 2011;129:687–94.
12. Parikh H, Deng Z, Yeager M, Boland J, Matthews C, Jia J, et al. A comprehensive resequence analysis of the KLK15-KLK3-KLK2 locus on chromosome 19q13.33. Hum Genet 2010;127:91–9.
13. Klein RJ, Hallden C, Cronin AM, Ploner A, Wiklund F, Bjartell AS, et al. Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer. Cancer Prev Res (Phila) 2010;3:611–9.
14. Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, et al. Prostate Testing for Cancer and Treatment (ProtecT) feasibility study. Health Technol Assess 2003;7:1–88.
15. Collin SM, Metcalfe C, Zuccolo L, Lewis SJ, Chen L, Cox A, et al. Association of folate-pathway gene polymorphisms with the risk of prostate cancer: a population-based nested case-control study, systematic review, and meta-analysis. Cancer Epidemiol Biomarkers Prev 2009;18:2528–39.
16. Chen L, Davey SG, Evans DM, Cox A, Lawlor DA, Donovan J, et al. Genetic variants in the vitamin D receptor are associated with advanced prostate cancer at diagnosis: findings from the Prostate Testing for Cancer and Treatment study and a systematic review. Cancer Epidemiol Biomarkers Prev 2009;18:2874–81.
17. Lamph SA, Sturgeon CM, White PAE, Price CP, Halloran SP. Evaluation report: total PSA assays. Centre for Evidence-based Purchasing (CEP), Purchasing and Supply Agency, National Health System. February 2009. CEP09004.
18. Guthrie PA, Gaunt TR, Abdollahi MR, Rodriguez S, Lawlor DA, Smith GD, Day IN. Amplification ratio control system for copy number variation genotyping. Nucleic Acids Res 2011;39:e54.
19. Atalay AC, Karaman MI, Guney S, Dalkilic A, Muslumanoglu AY, Ergenekon E. Age-specific PSA reference ranges in a group of non-urologic patients. Int Urol Nephrol 1998;30:587–91.
20. David A, Mabjeesh N, Azar I, Biton S, Engel S, Bernstein J, et al. Unusual alternative splicing within the human kallikrein genes KLK2 and KLK3 gives rise to novel prostate-specific proteins. J Biol Chem 2002;277:18084–90.
21. Michael IP, Kurlender L, Memari N, Yousef GM, Du D, Grass L, et al. Intron retention: a common splicing event within the human kallikrein gene family. Clin Chem 2005;51:506–15.
22. Kralovicova J, Gaunt TR, Rodriguez S, Wood PJ, Day IN, Vorechovsky I. Variants in the human insulin gene that affect pre-mRNA splicing: Is −23*Hph*I a functional single nucleotide polymorphism at IDDM2? Diabetes 2006;55:260–4.
23. Rodriguez S, Gaunt TR, Vorechovsky I, Kralovicova J, Wood PJ, Day IN. Comment on: Marchand and Polychronakos (2007) evaluation of polymorphic splicing in the mechanism of the association of the insulin gene with diabetes: Diabetes 2007;

56:709–713. Diabetes;56:e16.
24. Day IN, Rodriguez S, Kralovicova J, Wood PJ, Vorechovsky I, Gaunt TR. Questioning INS VNTR role in obesity and diabetes: subclasses tag IGF2-INS-TH haplotypes; and −23HphI as a STEP (splicing and translational efficiency polymorphism). Physiol Genomics 2006;28:113.
25. Raistrick CA, Day IN, Gaunt TR. Genome-wide data-mining of candidate human splice translational efficiency polymorphisms (STEPs) and an online database. PLoS One 2010;5:e13340.
26. McGee TL, Devoto M, Ott J, Berson EL, Dryja TP. Evidence that the penetrance of mutations at the RP11 locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous RP11 allele. Am J Hum Genet 1997;61:1059–66.
27. De AG, Rittenhouse HG, Mikolajczyk SD, Blair SL, Semjonow A. Twenty years of PSA: from prostate antigen to tumor marker. Rev Urol 2007;9:113–23.