DOI 10.1515/hsz-2012-0211 — Biol. Chem. 2012; 393(12): 1365–1390

Review

Jyotsna Batra*, Tracy O'Mara, Radhika Patnala, Felicity Lose and Judith A. Clements*

Genetic polymorphisms in the human tissue *kallikrein (KLK)* locus and their implication in various malignant and non-malignant diseases

Abstract: The *Kallikrein* (*KLK*) gene locus encodes a family of serine proteases and is the largest contiguous cluster of protease-encoding genes attributed an evolutionary age of 330 million years. The *KLK* locus has been implicated as a high susceptibility risk loci in numerous cancer studies through the last decade. The *KLK3* gene already has established clinical relevance as a biomarker in prostate cancer prognosis through its encoded protein, prostate-specific antigen. Data mined through genome-wide association studies (GWAS) and next-generation sequencing point to many important candidate single nucleotide polymorphisms (SNPs) in *KLK3* and other *KLK* genes. SNPs in the *KLK* locus have been found to be associated with several diseases including cancer, hypertension, cardiovascular disease and atopic dermatitis. Moreover, introducing a model incorporating SNPs to improve the efficiency of prostate-specific antigen in detecting malignant states of prostate cancer has been recently suggested. Establishing the functional relevance of these newly-discovered SNPs, and their interactions with each other, through *in silico* investigations followed by experimental validation, can accelerate the discovery of diagnostic and prognostic biomarkers. In this review, we discuss the various genetic association studies on the *KLK* loci identified either through candidate gene association studies or at the GWAS and post-GWAS front to aid researchers in streamlining their search for the most significant, relevant and therapeutically promising candidate *KLK* gene and/or SNP for future investigations.

Keywords: cancer; genome-wide association study; high-risk variants; imputation; kallikrein; single nucleotide polymorphism; tag single nucleotide polymorphism (SNP).

***Corresponding authors: Jyotsna Batra and Judith A. Clements,** Australian Prostate Cancer Research Centre-Queensland and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia, e-mail: jyotsna.batra@qut.edu.au; j.clements@qut.edu.au

Tracy O'Mara: Australian Prostate Cancer Research Centre-Queensland and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia; and Molecular Cancer Epidemiology, Queensland Institute of Medical Research, Brisbane, Queensland, Australia

Radhika Patnala: Australian Prostate Cancer Research Centre-Queensland and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

Felicity Lose: Molecular Cancer Epidemiology, Queensland Institute of Medical Research, Brisbane, Queensland, Australia

Introduction

Human genomic DNA is interspersed with many inter-individual differences. These approximately 3 million variations, also called polymorphisms, are estimated to span around 0.1% of the human genome (The International HapMap Consortium, 2007). The single nucleotide polymorphisms (SNPs) representing approximately 90% of all sequence variations is the most common type of variation in the human genome (Collins et al., 1998) with a frequency of >1% in a given population. However, some researchers distinguish between 'polymorphic SNPs' and 'common SNPs' with a minor allele frequency of at least 5% in the population (Brookes, 1999; Kruglyak and Nickerson, 2001; Ladiges et al., 2004).

Recent times have witnessed numerous candidate gene studies, which are restricted in the genes considered and the number of study subjects, and then whole-genome association studies (GWAS), making use of SNP arrays and elucidating various previously unknown disease-associated genes because of their unbiased advantage (Marian, 2012; Kote-Jarai et al., 2011b). The GWAS technique is constantly evolving, with efforts being made towards overcoming previous drawbacks, such as population stratification, false positives and negatives, and replication difficulties. Moreover, recent access to next-generation sequencing platforms has revolutionized investigation in the field of

genetic biomarkers by sharply reducing the cost of whole human-genome sequencing (Mardis, 2008a,b; Schuster, 2008; Metzker, 2010; Davey et al., 2011). The ultimate goals of these studies are to get a better understanding of the molecular mechanisms underlying a disease and to provide the foundation for the development of sensitive and readily applicable lab-based screening tools. Such tools will be useful for clinical diagnosis and monitoring of treatment and prognosis, and will thus assist clinicians to obtain a more accurate and reliable assessment of metastatic disease and hopefully support clinical decision-making.

The past few years have also seen an expansion in resources available to researchers, providing them with open access information and online data mining from the extensive data accumulated through various international collaborative efforts. For example, the HapMap project provides a comprehensive SNP database, containing information on linked genomic regions (Olivier, 2003). The 1000 Genomes Project (Siva, 2008; Pennisi, 2010) is an international collaboration to sequence the genomes of a substantial number of individuals to provide a comprehensive resource on human genetic variation and their haplotype blocks. The 1000 Genomes project (http://www.1000genomes.org), which is nearing completion, has already exceeded by half the number of novel SNPs observed by the HapMap project, i.e., with an estimate of more than 5.9 million variant nucleotide positions in the human genome. A total of 68 300 non-synonymous SNPs were identified through the 1000 Genomes pilot project, 34 161 of which were found to be novel. A fraction of these variations had been associated with various diseases and assigned a biological role, which is one step closer to establishing clinical relevance (Pennisi, 2010). For example, a particular 7 SNP risk profile may aid in the management of *BRCA2* mutation carriers in breast cancer (Antoniou et al., 2010), and a combination of several validated 'low-risk' SNP markers has been proposed to be useful in breast cancer and prostate cancer risk prediction (Pharoah et al., 2008; Zheng et al., 2008).

The protein-based indicator prostate-specific antigen (PSA), encoded by the *KLK3* gene at the *kallikrein (KLK)* locus, is a well known and widely used oncogenic biomarker to diagnose and monitor prostate cancer progression after metastasis and treatment (Tan Olivia et al., 2006; Lawrence et al., 2010). It has accrued many critics since its use, however, mostly due to its non specificity and accuracy in establishing aggressive prostate cancer prognosis over possible benign tumors, and on the identification of set threshold levels to determine the need for invasive techniques such as biopsies (Prensner et al., 2012), thus calling for a timely investigation to fine-tune the efficiency of PSA as a prostate cancer biomarker, possibly through the detailed investigation of the *kallikrein* gene locus and the genetic variants at this locus, and to uncover the functional aspects of this disease-associated region.

The *KLK* locus, which features among the seven highly-ranked susceptibility loci in a multi-stage prostate cancer GWAS (Kote-Jarai et al., 2008, 2011a), is clustered in a tandem array of approximately 300 kilobases (kb) on chromosome 19q13.4, and contains the largest cluster of 15 homologous protease genes (Lawrence et al., 2010) said to have evolutionarily emerged 330 million years ago (Clements, 2008; Pavlopoulou et al., 2010). With the recent imperative given to research on the *KLK* gene locus, many significant gene – disease associations have been established and previously unknown roles of KLK proteins elaborated. Since the initial observation that the three classical *kallikreins* – human kallikrein 1 (hk1/KLK1), human kallikrein 2 (hk2/KLK2) and PSA – are localized in this region, an additional 12 newly-discovered *kallikreins* have been mapped to the KLK locus (Paliouras and Diamandis, 2006). Their secreted extracellular nature makes the KLK proteins potential targets as biomarkers that can easily be analyzed by the well-established enzyme-linked immunosorbent assay (ELISA) method. Apart from their importance as potential cancer biomarkers, their imprint has also been made in diseases like diabetes, skin disorders and neurodegenerative diseases (Paliouras and Diamandis, 2006).

We have reviewed the genetic architecture, genetic isoforms, function and the role of kallikreins in disease metastasis in previous articles (Tan Olivia et al., 2006; Lawrence et al., 2007, 2010). To the best of our knowledge, this article catalogs most of the identified SNPs within the KLK locus and reviews the role of these polymorphisms in various malignant and non-malignant diseases established in the context of gene-association studies conducted to date, with the objective of providing the readership with a consolidated resource on relevant SNPs in the *KLK* locus for future disease association studies.

KLK SNP data-mining from SNPdb and 1000 Genomes

We assessed all the SNPs recorded within ±10 kb of the *KLK* locus mapped in Genome Build GRCh37/hg19 (chr19:51312404..51587502), from the most popular public database of SNPs, the National Center for Biotechnology Information's dbSNP (Sherry et al., 2001) build 132, using the UCSC web-browser (http://genome.ucsc.edu/). From a total of 4331 polymorphisms identified within the *KLK* locus, 3420 (73.4%) were found to be SNPs and 911 (26.6%) were

insertion/deletion polymorphisms (indels) and/or mixed type. A similar study was published by Goard et al., in 2007 using custom designed tools 'ParSNPs' and 'LocusAnnotator', which generated a catalog of 1856 polymorphisms of which 1023 were validated (Goard et al., 2007). While in our analysis a total of 2627 out of the 4331 polymorphisms were found to be validated (Nov, 2011) either by frequency, two-hit, submitter, cluster or by HapMap, which includes 2535 SNPs and 92 indels. (dbSNP uses certain validation methods to identify SNPs for their relevance. By frequency, where at least one submitted SNP in the cluster should have frequency data submitted; by cluster, which has two submissions with at least one submission assayed with a non-computational method; by submitter, with at least one submitter validated by independent assay. By two-hit/two-allele signifies that all alleles have been observed in at least two chromosomes). Five-hundred-seventy-four SNPs out of 2627 have recently been discovered from the 1000 Genomes database (Siva, 2008; Pennisi, 2010). A total of 2150 SNPs were shown to have a minor allele frequency (MAF) >1% in the European population.

Our further analysis was restricted to validated polymorphisms only in an attempt to avoid analysis of false-positive records in the dbSNP database that might have arisen due to sequencing artifacts. Based on the nature of their alleles, 1678 (66.2%) out of the 2535 validated SNPs in the *KLK* locus (66.2%) were C→T transitions (or G→A on the opposite strand), and 464 (18.3%) were C→A (G→T on the opposite strand), while 234 (9.2%) and 148 (5.8%) were C→G and T→A transversions, respectively. The records for 11 SNPs, which were reported to constitute more than two alleles, were not classified in the above groups. Considering the functional effects of SNPs based on their position to the nearest gene, and as defined by the UCSC browser, the most prevalent class annotations corresponded to those lying in non-coding intronic class (678/2535; 26.7%), untranslated region (113, 4.5%) and gene locus region polymorphisms (241, 9.5%, present near the 3' or 5' of the gene). A smaller proportion of polymorphisms were associated with coding regions in the *KLK* locus (97/2535; 3.8%), with 35% synonymous, 62% missense and 3% nonsense changes. The remaining polymorphisms not attributed to any position class by dbSNP (1406/2535; 55.4%) may refer to either intergenic or unannotated polymorphisms (Figure 1).

Functional *KLK* SNPs: *in silico* annotation

Functional SNPs are those polymorphisms that, depending on their position with respect to a protein coding region and/or a regulatory site, can affect the gene function and thus have a tendency to alter the cellular processes and functioning of the cell. Figure 2 details various web-based tools that can be used for the detailed *in silico* prediction and analysis of potential functional roles of the *KLK* SNPs. To this end, data for 1404 out of 2535 SNPs were retrieved using 'FuncPred' (http://snpinfo.niehs.nih.gov/snpfunc.htm) from the SNPinfo web-server (http://manticore.niehs.nih.gov/snpfunc.htm), which assesses multiple functional prediction programs as well as calculating the regulatory potential score and conservation scores of SNPs (i.e., protein stability, splicing regulation, transcriptional regulation and post-translational modification), but lacking updated SNPs information from the 1000 Genomes project. Fourteen SNPs (rs11670728, rs12974899, rs12978483, rs2569522, rs2659056, rs28384475, rs3212811, rs3212840, rs3212846, rs3760739, rs58876874, rs7252452, rs3212850 and rs3745541) were predicted to alter a transcription-factor binding site. Nine SNPs (rs1624358, rs16988799, rs2736433, rs28384475, rs35192866,

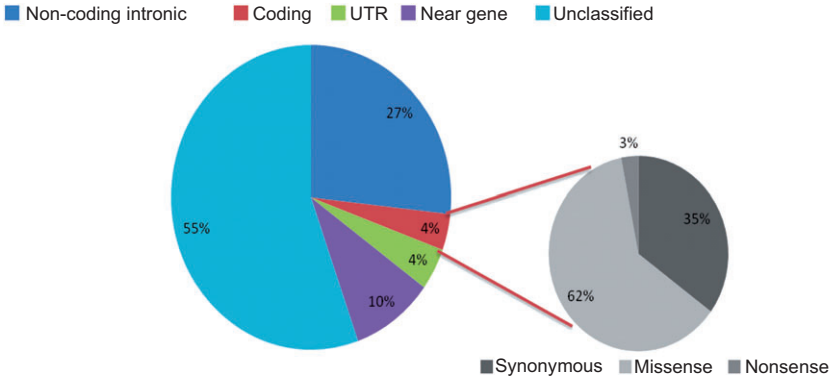

Figure 1 Position-based class annotations associated with validated *KLK* polymorphisms in dbSNP as downloaded from the UCSC browser. Note: As multiple transcript variants are known for each *KLK* gene, the functional class annotations may vary based upon the transcript under consideration. In the current annotation, single nucleotide polymorphisms have been labeled on a preferential basis, e.g., coding region preferred over untranslated region (UTR), followed by intronic and near gene locations.

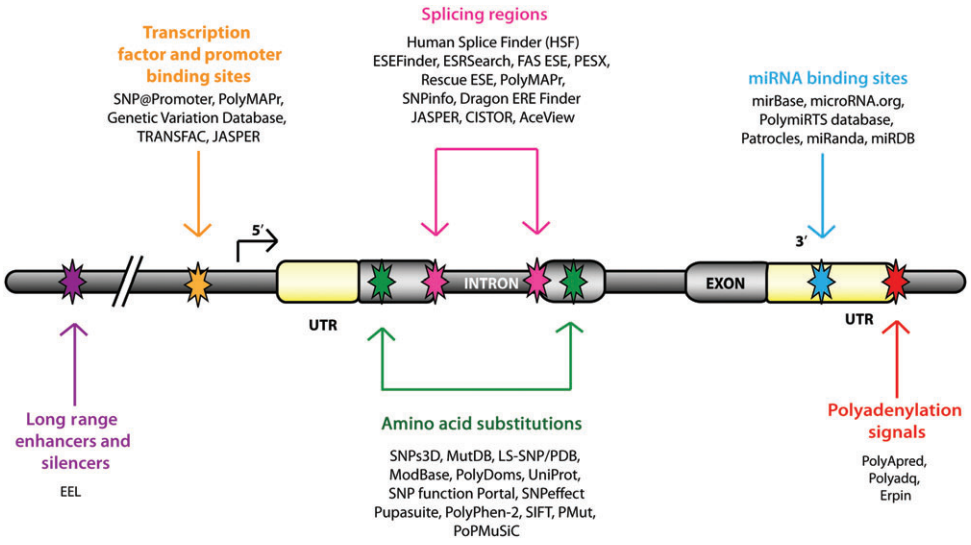

Figure 2 An overview of some of the web-based tools used are represented relative to the single nucleotide polymorphism location in a gene (Figure adapted from Lee et al., 2009b).
Each single nucleotide polymorphism should be further examined for its functional effect with respect to each category (i.e., protein coding, splicing regulation, transcriptional regulation and post-translation) using a series of algorithms.

rs61752567, rs7259651, rs10403407 and rs2659094) were predicted to alter the exonic splicing enhancer sites and disturb splicing regulation, while six might abolish the splicing domain itself. The PolyPhen (http://coot.embl.de/PolyPhen/) tool used by FuncPred (Adzhubei et al., 2010) estimated the structural and functional impact of an amino acid substitution and predicted five SNPs to be deleterious (rs1048328, rs198977, rs5515, rs10422897 and rs183854). The SNPs3D software (http://www.snps3d.org; Yue et al., 2006) was used to predict any deleterious effect of SNPs on protein function, which it does by making predictions based on the estimated impact of a non-synonymous SNP on protein stability, or considering conservation of the given amino acid within a protein family. Eleven SNPs (rs17632542, rs5515, rs198977, rs6072, 3733402, rs3733402, rs4253325, rs4253379, rs2569527, rs1048328 and rs183854) were predicted to be affirmative in their deleterious effect on protein functionality by SNPs3D. Twenty-four SNPs were predicted to have a high conservation score of >0.4 and three SNPs (rs7245858, rs2691209 and rs16989073) to have a score of 1.

Another important class of functionally relevant SNPs are polymorphisms present at or near the microRNA binding sites of functional genes (miRSNPs) with the potential to interfere with miRNA function, thus affecting gene expression (Pelletier and Weidhaas, 2010). Among the *KLK* SNPs, four (rs10426, rs2691258, rs58682039 and rs61269009) were predicted to alter miRNA-binding sites, as predicted by Miranda (http://www.microrna.org/microrna/home.do), while the Sanger method (http://www.mirbase.org/) predicted nine miRSNPs (rs2411334, rs2569735, rs2659092, rs268883, rs4846, rs9524, rs12151211, rs1654555 and rs2232539) within the KLK locus.

Functional *KLK* SNPs: experimental validation

Apart from the various *in silico* methods mentioned above, naturally-occurring polymorphisms identified in many human *KLK* genes (e.g., *KLK1, KLK2, KLK3* and *KLK12*) have had their functional role established through laboratory investigation, as summarized in Table 1.

KLK1 activity has previously been reported to decrease in people with primary hypertension and to be partly inherited, which led to the investigation of the association of *KLK1* exon 3 non-synonymous polymorphism Arg77His (dbSNP ID rs5515) with urinary *KLK1* activity in hypertensive individuals (Slim et al., 2002). Out of the 66 patients analyzed, five who were heterozygous for the Arg77His polymorphism were seen to harbor statistically significant down-regulation of urinary KLK1 activity. *In vitro* functional analysis of the activity of the wild-type or polymorphic KLK1 protein confirmed this decreased activity in the presence of the histidine residue, and modeling using crystallographic data suggested this residue may alter substrate binding. The same laboratory performed a follow-up study of this finding in normotensive subjects and confirmed the reduced activity in those carrying the histidine allele (Slim et al., 2002). In

Gene	SNP and location	Genotyping method	Associated disease (sample used)	Molecular causation	References
KLK1	*KLK1* exon 3 non-synonymous polymorphism Arg77His (dbSNP ID rs5515)	Direct sequencing, PCR-RFLP	Hypertension (plasma and/or urine)	Lowered gene expression recorded. Histidine residue alters substrate binding	(Slim et al., 2002)
		Direct sequencing	Cardiovascular disease (blood)	Inappropriate remodeling of the brachial artery	(Azizi et al., 2005)
	KLK1 promoter region (between -133 and -121 with respect to the transcription initiation site)	PCR and DNA sequencing	Hypertension (blood)	Modulation of promoter activity	(Song et al., 1997)
KLK2 (encoding kallikrein-related peptidase 2; hK2/KLK2)	rs198977, which substitutes Arg250 for a tryptophan residue	PCR and DNA sequencing	Prostate cancer susceptibility (prostatic tissue and blood leukocytes)	Lack of trypsin-like activity by the polymorphic *KLK2* protein. Increased serum hK2/KLK2 levels	(Herrala et al., 1997)
KLK3 (encoding prostate specific antigen; PSA)	rs266882 SNP, within ARE1 of *KLK3*	PCR-RFLP	Prostate cancer susceptibility (peripheral blood leukocytes)	Differential binding of ARE1 to the androgen receptor. Increased serum PSA levels	(Lai et al., 2007)
KLK7	*KLK7* 3' untranslated region AACC insertion	PCR-RFLP	Atopic dermatitis (whole blood)	Could alter translational efficiency through post-translational modifications (needs experimental validation)	(Vasilopoulos et al., 2011)
KLK12	*KLK12* intronic c.457+2T>C polymorphism	PCR-SSCP and DNA sequencing	Primary gastric cancer (non-cancerous gastric tissue)	Splicing abnormality	(Shinmura et al., 2004)

Table 1 Experimentally validated single nucleotide polymorphisms in *KLK* genes (*KLK1*, *KLK2*, *KLK3*, *KLK7* and *KLK12*).
PCR, polymerase chain reaction; RFLP, restriction fragments length polymorphism; SSCP, single-strand conformation polymorphism; SNP, single nucleotide polymorphism.

addition, the individuals heterozygous for Arg77His also showed inappropriate remodeling of the brachial artery, which suggested implications for cardiovascular disease (Azizi et al., 2005). These studies support the finding of SNP rs5515 in the *in silico* analysis mentioned above, which also predicted it to possibly be deleterious. Another *KLK1* coding SNP, Gln145Glu (rs5516), was not shown to affect urinary *KLK1* activity in this study (Slim et al., 2002). Incidentally, a functional analysis in 1997 reported SNPs/fragment length polymorphism in the promoter region of *KLK1* (between -133 and -121 with respect to the transcription initiation site) to be associated with decreased *KLK1* gene expression (Table 1; Song et al., 1997).

KLK2 displays a common coding region polymorphism that substitutes Arg250 for a tryptophan residue (rs198977). Experimental analysis using recombinant KLK2 in insect cells revealed a lack of trypsin-like activity by the polymorphic KLK2 protein (Table 1; Herrala et al., 1997). The rs198977 SNP has also been shown to be associated with significantly decreased levels of serum KLK2 in two large studies (Nam et al., 2003; Klein et al., 2010), and our *in silico* analyses in the last subsection indicated a possibly damaging/deleterious effect. This SNP has been the subject of much investigation in relation to prostate cancer risk and will be discussed later in this review.

The *KLK3* rs266882 SNP is one of the most interesting functional SNPs in the context of prostate cancer susceptibility. Residing within one of the androgen response elements (AREs) of *KLK3*, ARE1, various studies have reported its association with increased serum PSA levels (Xue et al., 2001; Medeiros et al., 2002; Xu et al., 2002; Rao et al., 2003; Schatzl et al., 2005). Functional studies on rs266882 attribute *KLK3* gene expression alteration to differential binding of ARE1 to the androgen receptor (Table 1) and enhanced transcriptional response to androgens. They also show increased prostate cancer susceptibility in the presence of A/A genotype (three-fold risk) and A/G genotype (2.4-fold risk; Lai et al., 2007).

Vasilopoulos et al. investigated the *KLK7* 3' untranslated region AACCins5874 insertion polymorphism for functional effects on *KLK7* expression and observed an increased expression in the presence of the insertion (Table 1; Vasilopoulos et al., 2011), although no convincing association of this polymorphism with disease has been established to date.

The *KLK12* intronic c.457+2T>C polymorphism, i.e., a T to C substitution in the second nucleotide of intron 2, was claimed to be associated with a splicing abnormality (Table 1), with the expression of the human *KLK12* classical mRNA and the protein (hK12/KLK12) corresponding to the putative serine protease being absent in individuals with a c.457+2C/C genotype but not in individuals with the T/T or T/C genotypes (Shinmura et al., 2004). However, the small number (n=22) of samples considered in the analysis cells for replication in a larger set for results to be validated and conclusively confirmed.

Additional experimental studies are required to validate the functional role of the SNPs shortlisted by the *in silico* methods mentioned above.

Genetic architecture of the *KLK* locus: haplotypes and tagging

Alleles of SNPs in close physical proximity to each other are often correlated, and can be represented as haplotypes. Linkage disequilibrium (LD) is the occurrence of some combinations of alleles within a population more or less often than would be expected from a random formation of haplotypes from alleles (Devlin and Risch, 1995). LD is represented by an r^2 or a D' value, which is calculated based on the difference between observed and expected allelic frequencies (assuming random distributions). Geneticists commonly use the threshold of $r^2 > 0.8$ to measure which SNPs are in LD with other SNPs. To draw the LD map of the *KLK* locus, we used the HapMap Public Release 27, Build 36 (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap27_B36/#search). All database SNPs within the Centre d'Etude du Polymorphisme Humain (CEPH) population (Utah residents with ancestry from northern and western Europe) were plotted using the Haploview 4.2 software (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/downloads; Barrett et al., 2005).

A total of 262 SNPs were genotyped in HapMap for a total of 205 individuals; 19 SNPs had a frequency of <0.01, and so were not included in haplotype analysis. A total of 33 LD blocks were identified for the 15 *KLK* genes, as shown in Figure 3. A tag SNP is a representative of a group of SNPs in a region of the genome with high LD. Tag SNPs are useful in whole-genome SNP association studies (described in detail later) as it is then possible to confidently predict genotypes for a number of different SNPs without assessing every SNP in a chromosomal region. The number of tagging SNPs within the KLK loci required to represent those HapMap SNPs with a frequency of <0.01 was determined using the 'Tagger' functionality with Haploview by means of a pairwise analysis of LD with an r^2 threshold of 0.8. Genotyping of 156 representative tagging SNPs is required to cover the 243 *KLK* variants based on

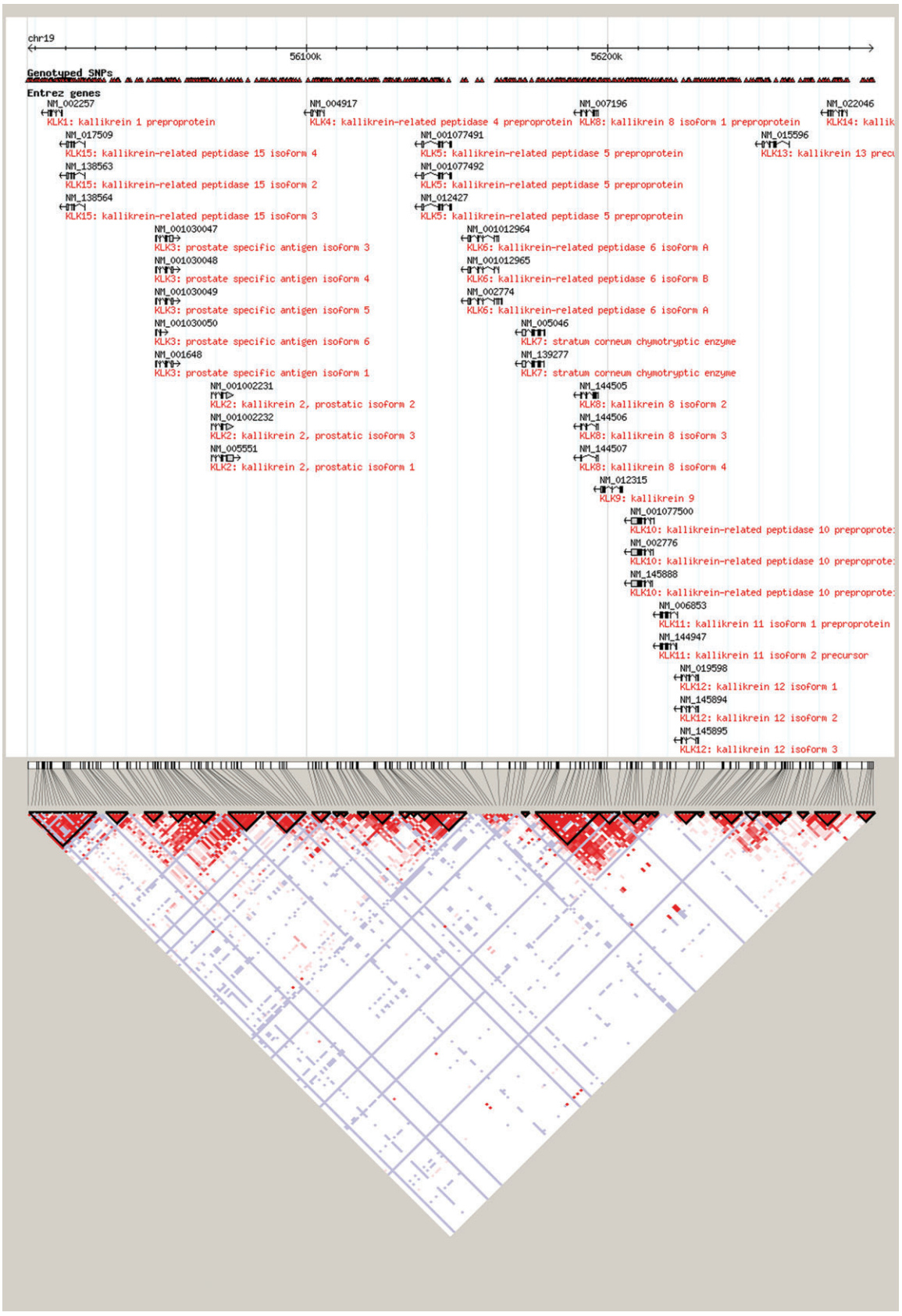

Figure 3 Linkage disequilibrium map of the *KLK* locus plotted using Haploview v4.2.
Data from the HapMap database European population was used.

a pairwise analysis of LD with an r^2 threshold of 0.8. As discussed above, the 1000 Genomes project is identifying many more genetic variants than those included in the HapMap database, and hence a vastly refined LD map of the *KLK* locus can be produced when the data are released in a user-friendly format.

Recently, Parikh et al. (2010) generated a comprehensive LD map of common SNPs and indels through deep sequencing analysis of a 56 kb region covering the KLK15–*KLK3*–*KLK2* genes (chr 19: 56,019,829-56,076,043 bp; NCBI Build 36.3) in 78 unrelated individuals of European ancestry. A total of 555 polymorphic loci were identified, including 116 novel SNPs and 182 novel indels. Based on tagging analysis, 144 SNPs are necessary to tag the region at an r^2 threshold of 0.8 and MAF of 1% or higher, while 86 loci are required to tag all SNPs with a MAF >5%. Further, these sequence data augment coverage of this region by only 35% and 78% compared to variants in dbSNP and HapMap, respectively. Thus, additional studies on LD mapping and tag SNP selection based on deep sequencing data are highly recommended.

Genetic association studies and *KLK* SNPs

Numerous studies have been performed to investigate SNPs found in the *KLK* genes for their malignant and non-malignant disease associations. The majority of SNP association studies have focused on their effects on cancer risk, particularly prostate cancer. Presumably, this was originally because of the identification of PSA/*KLK3* in the etiology of this disease, and more recently because of results from large genome-wide studies implicating this locus in prostate cancer risk. Below we review the various SNP association studies carried out to determine disease susceptibility and the methods used to perform these analyses. The genotyping for the association studies was conducted on blood genomic DNA unless otherwise specified.

Association studies on low risk variants in *KLK* genes

Inherited genetic variants can be segregated into two categories: rare high-risk genetic variants (or mutations) and common low-risk genetic variants, such as SNPs. This review subsequently focuses on common, low-risk genetic variants or SNPs of the *KLKs* and the genetic approaches used to identify these. While these SNPs have much weaker effects than high-risk genetic mutations, they are common and may have a large population attributable risk (>5%) and together could contribute to a complex disease phenotype (Pestell, 2008). A complete and detailed list of SNPs analyzed in gene-association studies in previous research initiatives, including those that failed to reach statistical significance, has been included in the supplementary tables.

Pre-GWAS KLK candidate gene-association studies

Prior to 2007, the candidate-gene approach was the predominant method used to explore inherited low-risk genetic variants. This approach is based on *a priori* knowledge about the gene(s) of interest in the pathogenesis of the phenotype and involves an examination of a relatively small number of genetic variants (between 1 and 100 SNPs; Savage, 2008). The approach has led to the identification of a number of SNPs that may influence the risk of hormone-related cancers, particularly prostate cancer, with instances of ovarian cancer and breast cancer, as listed in Table 2.

The most frequently analyzed *KLK* SNP before 2008 was rs266882 (G-158A) in the *KLK3* locus; a significant association with prostate cancer risk for this SNP was originally reported by Xue et al., in 2000 (Table 2). Numerous subsequent studies have revealed conflicting results. Some have reported an association with the same allele as the original study (Gsur et al., 2002; Medeiros et al., 2002; Cicek et al., 2005; Lai et al., 2007), while others have reported either a significant association in the opposite direction (Chiang et al., 2004; Binnie et al., 2005) or no association for this SNP (Wang et al., 2003; Salinas et al., 2005; Mononen et al., 2006; Severi et al., 2006; Cunningham et al., 2007; Pal et al., 2007; Penney et al., 2011). A meta-analysis of all studies published up until 2008 reported no evidence of association with prostate cancer risk overall (Supplementary Table 1; Jesser et al., 2008).

Currently, significant associations have been reported for SNPs located within *KLK2*, *KLK3*, *KLK4*, *KLK10* and *KLK15* with breast cancer, ovarian cancer survival and prostate cancer. Genes and significant SNPs of *KLK2* (rs198977, rs2664155 and rs1506684), *KLK4* (rs806019), *KLK10* (rs3745535) and *KLK15* (rs266851) found to be associated with cancer risk or prognosis are detailed in Table 2. Information on subsequent studies, especially of those SNPs that failed to reach significance, can be found in Supplementary Table 1.

Gene	SNP	Alleles	Genotyping platform	Disease (sample used)	Putative functional role*	Cases	Controls	Association?	Risk estimates (95% CI)	References
KLK2	rs198977 (C792T)	C/T	PCR and DNA sequencing	PrCa (blood)	miRNA binding non-syn, damaging by polyphen	617	671	Yes	TT: OR 2.13 (1.3–3.5); p=0.004	(Nam et al., 2003)
			Sequenom MassARRAY	PrCa (blood)		996	1092	Yes	TT: OR 1.49 (1.0–2.2); p=0.04	(Nam et al., 2005)
			PCR-RFLP	PrCa (peripheral venous blood)		254	168, men with BPH	Yes	C allele: cases 82.1%, controls 74.7%; p=0.010	(Chiang et al., 2005)
			Sequenom MassARRAY DNA Sequencing	PrCa (blood)		645	606	Yes	TT/CT: OR 1.3 (1.1–1.6); p=0.05	(Nam et al., 2006)
			PCR-RFLP	PrCa (peripheral blood leukocytes)		135	142	Yes	CC: OR 2.78 (1.09–7.06); p=0.031	(Mittal et al., 2007)
			GoldenGate™ MassArray (Illumina)	BrCa (Peripheral blood)		117	194	Yes	T allele: OR 0.56 (0.34–0.83); p=0.0059	(Lee et al., 2009a,b)
KLK2	rs2664155	G/A	Sequenom MassARRAY, DNA Sequencing	PrCa (blood)	TFBS	645	606	Yes	AG/AA: OR 1.4 (1.2–1.8); p=0.002	(Nam et al., 2006)
KLK2	rs1506684	C/T	SNPlex genotyping system (OLA/PCR)	PrCa	TFBS	596	567	Yes	T allele: 47.2% (controls) vs. 42.4% (cases); p=0.041	(Pal et al., 2007)
KLK3	rs266882	G/A	PCR-RFLP	PrCa (normal seminal vesicle tissue)	TFBS	100, advanced cases	100	Yes	GG: OR 2.90 (1.24–6.78)	(Xue et al., 2000)
	G-158A		PCR-RFLP	PrCa (venous blood)		151	127	Yes	A allele: 63.3% vs. 48.8%; p=0.009	(Medeiros et al., 2002)
								Yes, with early onset PrCa	AA: OR 2.92 (1.10–7.86) p=0.013	
								Yes, with circulating tumour cells	AA: increased circulating tumour cells (p=0.018)	
			PCR-RFLP	PrCa (mononuclear cells from blood)		132	149	Yes	AG+GG: OR 0.63 (0.39–0.99); p=0.048	(Gsur et al., 2002)
								Yes, with increased Gleason score	GG: OR 2.29 (1.06–4.94); p=0.034	
			PCR-RFLP	PrCa (blood)		122	84, BPH patients	Yes	G allele: 87.3% vs. 77.4%; p=0.008	(Chiang et al., 2004)
								Yes	GG: OR 2.27; p=0.008	
								Yes, with larger tumour volume	GG: larger tumour volume (p=0.013)	
								Yes, with increased extracapsular extension	GG: increased extracapsular extension (p=0.036)	

Table 2 (Continued)

Gene	SNP	Alleles	Genotyping platform	Disease (sample used)	Putative functional role*	Cases	Controls	Association?	Risk estimates (95% CI)	References
			PCR-RFLP	PrCa (blood)		439	479	Yes	GG: OR 2.71 (1.06–6.94); p=0.04	(Cicek et al., 2005)
			PCR-RFLP	PrCa (blood)		100	67	Yes	GG: 30% cancer vs. 16% controls (p=0.025)	(Binnie et al., 2005)
			Sequenom MassARRAY	PrCa (peripheral blood leukocytes)		209	223	Yes	AA/AG: OR 2.61 (1.37–4.96) p=0.004	(Lai et al., 2007)
								Yes, with Gleason score	GG: OR 6.23 (2.29–16.98); p<0.01	
KLK3	rs11084033	C/A	Sequenom MassARRAY	OvCa (blood)	TFBS	304	–	Yes, with survival	AA: HR 2.12 (1.08–4.15); p=0.04	(O'Mara et al., 2011)
KLK3	rs11575894	A/AA		BrCa		101	–	Yes, with less aggressive cancer	Less aggressive cancer	(Yang et al., 2002)
KLK3	rs925013	G/A	PCR-RFLP	PrCa (blood)	TFBS	439	479	No	Not significant	(Cicek et al., 2005)
	(G-4643A)		Sequenom MassARRAY	PrCa (blood)		821	734	Yes	G allele: OR 1.4 (1.1–1.7); p=0.001	(Severi et al., 2006)
KLK3	rs266849	A/G	SNPlex Genotyping System (OLA/PCR)	PrCa		596	567	Yes	G allele: 20.7% (controls) vs. 15.7% (cases); p=0.006	(Pal et al., 2007)
KLK3	rs1058205	T/C	SNPlex Genotyping System (OLA/PCR)	PrCa	miRNA binding site	596	567	Yes	C allele: 20.2% (controls) vs. 14.2% (cases); p=0.001	(Pal et al., 2007)
KLK3	rs266870	C/T	SNPlex Genotyping System (OLA/PCR)	PrCa		596	567	Yes	T-allele: 52.3% (controls) vs. 45.4% (cases), p=0.004	(Pal et al., 2007)
KLK3	rs2659122	A/G	SNPlex Genotyping System (OLA/PCR)	PrCa	miRNA binding site	596	567	Yes	G allele: 28.1% (controls) vs. 23.9% (cases); p=0.041	(Pal et al., 2007)
KLK4	rs806019	C/G	GoldenGate™ assay (Illumina)	BrCa (peripheral blood)		117	194	Yes	G allele: OR 0.53 (0.33–0.85); p=0.0068	(Lee et al., 2009a,b)
KLK10	rs3745535	G/T	DNA Sequencing	PrCa (tissue and whole blood)	Splicing (ESE or ESS), non-syn	49	52	Yes	GG: cases 26% vv. controls 50%; p=0.027	(Bharaj et al., 2002)
KLK15	rs266851	C/T	Sequenom MassARRAY	OvCa (blood)	TFBS	319, Australian cases	–	Yes	CT/TT: OR 1.42 (1.02–1.96); p=0.01	(Batra et al., 2011b)

Table 2 Summary of KLK candidate gene association studies performed in hormone-related cancers.
*As predicted by 'FuncPred' from the SNPinfo web-server (http://manticore.niehs.nih.gov/snpfunc.htm). BrCa, breast cancer; CI, confidence interval; GWAS, genome-wide association study; OvCa, ovarian cancer; OR, odds ratio; PCR, polymerase chain reaction; PrCa, prostate cancer; TCGA, The Cancer Genome Atlas; TFBS, transcription factor binding site; ESE, exonic splicing enhancer; ESS, exonic splicing silencer; non-syn, non-synonymous; OLA, oligonucleotide ligation assay.

Regarding the potential role of polymorphisms in non-cancerous diseases, *KLK1* has been the most intensively studied *KLK* gene, perhaps because it was the first *KLK* gene to be discovered. Of the 13 different non-malignant diseases and traits investigated in *KLKs* (Table 3), the role of *KLK1* in cardiovascular and kidney-related diseases has been the focus of the majority of studies. A candidate gene association study for the functional SNP rs5515 with risk of cardiovascular traits/disease did not report any association in the two small candidate SNP studies performed to date (Supplementary Table 2). Perhaps the most compelling finding arose from research into hypertension in the Chinese Han population, with several *KLK1* polymorphisms indicating an increased cardiovascular traits/disease risk (Table 3, Supplementary Table 2; Hua et al., 2005; Zhao et al., 2007). Other examples of disease associations include *KLK4* for AIH IIA1, *KLK7* for atopic dermatitis, and *KLK8* for intracranial aneurysm, details of which can be found in Table 3.

A recent small cohort study (cases=218 and controls=220) by Lee et al., performed pyrosequencing to genotype the SNP site (+255G>A, rs2664155) in intron 1 of the *KLK2* gene. A statistically significant association between rs2664155 and male infertility (OR=0.47, 95% CI=0.26–0.85, $p<0.05$) was reported (Lee and Lee, 2011), although no functional analysis towards this risk susceptibility has been carried out to date.

Unless performed in a very large sample size from a well-characterized population, the candidate-gene approach is prone to spurious results. This is evident by the number of follow-up studies performed that have attempted to validate positive reports of *KLK* polymorphism disease associations that have, however, failed to confirm original results (Supplementary Tables 1 and 2).

KLK genes in genome-wide association studies

Since around 2007, genetic epidemiology has been transformed by the availability of high-throughput genotyping methods designed to provide an unbiased survey on the effects of common genetic variants, called GWAS. GWAS are studies wherein research subjects are typed for a large number of genetic variants, typically between 300 000 and 1 000 000 polymorphisms, and the allele or genotype frequencies are evaluated for differences between groups (e.g., disease vs. non-disease groups). The advantage of GWAS is that they allow for a wide search of genetic variants associated with disease without having to specify a particular gene of interest, facilitating the mining of potentially novel variants (Wellcome Trust Case Control Consortium, 2007). Due to the massive number of joint statistical tests performed, however, there is a higher level of type-1 error (false-positives). Statistical corrections for multiple hypotheses testing are therefore essential and a $p<10^{-7}$ has been proposed as an appropriate significance level for evidence of a genome-wide association (Thomas et al., 2005). This means that large sample sizes are required for GWAS to ensure adequate statistical power to detect an association with small *p*-values. Since the advent of GWAS technology, highly statistically significant and robust associations with SNPs in over 230 diseases and traits have been successfully identified (Hindorff et al., 2011). The National Human Genome Research Institute maintains a catalog of published GWAS that can be accessed at http://www.genome.gov/gwastudies/.

A pioneering GWAS performed with 3268 cases and 3366 controls identified a prostate cancer susceptibility locus between *KLK2* and *KLK3*. The minor allele of an SNP (rs2735839) in this region was reported to confer a 1.2-fold decreased risk of prostate cancer (per allele odds ratio [OR] 0.83, 95% confidence interval [CI] =0.75–0.91; $p=1.5\times10^{-18}$; Table 4; Eeles et al., 2008). Following this discovery, there have been a large number of studies pursuing association studies of SNPs in the *KLK* region, examining both prostate cancer risk and prognostic features. The *KLK* gene associations found to be significant to prostate cancer susceptibility are summarized in Table 4, with non-significant associations mentioned in Supplementary Table 3. Relevant studies are discussed below.

The above GWAS-identified SNP, rs2735839, displayed a strong association with PSA levels (Table 4; Eeles et al., 2008). However, there has been some debate as to whether this SNP is truly associated with prostate cancer or simply relates to PSA expression levels, since male controls used for the stage 1 analysis were limited to those with clinically low PSA levels (<0.5 ng/ml; Ahn et al., 2008). Nevertheless, further investigation of rs2735839 in additional large case-control sample sets where controls were not screened for PSA levels did replicate the association (Kote-Jarai et al., 2008). There have been several studies performed to find more such risk alleles, assessing the association of SNPs in the *KLK2* and *KLK3* region with PSA and hK2/*KLK2* levels (Supplementary Table 4).

An evaluation study was carried out by Bensen et al. to determine SNP associations with prostate cancer aggressiveness in Afro-American and European-American men from the North Carolina-Louisiana Prostate Cancer Project. Genotyped DNA from blood and buccal cell samples demonstrated three SNPs in the KLK3 region (i.e., rs266870,

Outcome	Gene	Polymorphism/mutation	Genotyping method, functional effect	Ethnicity	Cases	Controls	Association	Risk estimates OR (95% CI)	References
Acute kidney injury*	*KLK1*	Methylation	Pyrosequencing (CpG methylation analysis) May alter expression	Caucasian/ African American	14	32	Methylation	Increased: p=0.011	(Kang et al., 2011)
Cerebral hemorrhage	*KLK1*	Polymorphisms: rs5516 (Gln145Glu), rs5517 (Lys186Glu)	SNaPshot and direct sequencing non-synonymous SNPs; may also alter isoform expression	Chinese Han	273	140	rs5517	Increased risk: p<0.05	(Zeng et al., 2010)
Hypertension	*KLK1*	Polymorphism: rs3212816, rs78093423 (-130 G_n)	PCR-RFLP allele-specific oligonucleotide (ASO) hybridization	Chinese Han	200	200	rs3212816	Increased risk: p<0.05	(Hua et al., 2005)
				Chinese Han	200	200	-130 G_s	Increased risk: p<0.05	
Hypertension	*KLK1*	Polymorphisms: rs5516 (Gln145Glu), rs5517 (Lys186Glu)	PCR-RFLP non-synonymous SNPs; may also alter isoform expression	Chinese Han	2411	2348	rs5517	Increased risk: 1.25 (1.16–1.46), p=0.007	(Zhao et al., 2007)
Hypertension-associated end-stage renal disease	*KLK1*	Polymorphism: rs78093423 (-130 G_n)	Semiautomated sequencing, direct DNA sequencing, manual sequencing; alters expression	African American	76	85	-130 G_{12}	Increased risk: p=0.003	(Yu et al., 2002)
Abdominal aortic aneurysm	*KLK1*	Polymorphism: rs5516 C>G (Gln145Glu)	Direct sequencing, non-synonymous SNP; may also alter isoform expression	Caucasian	755 small AAA, 79 large AAA	795	G allele	Borderline increased risk: 2.40 (0.98–5.88), p=0.056	(Biros et al., 2011)
Vesicoureteric reflux with renal progression in children	*KLK1*	Polymorphism: rs78093423 (-130 G_n)	PCR amplification-SSCP; alters expression	Taiwanese	28 (with progression), 120 (no progression)	170	-130 G_{12}	Increased risk: p=0.008	(Lee-Chen et al., 2004)
Amelogenesis imperfecta-hypomaturation type IIA1	*KLK4*	Mutation: g.2142 G>A (Trp153Stop)	DNA sequencing, truncated protein	Not reported	1 family	N/A	A allele	Causal	(Hart et al., 2004)
Amelogenesis imperfecta-hypomaturation type IIA1	*KLK4*	Mutation: g.2142 G>A (Trp153Stop)	DNA sequencing, truncated protein	Caucasian	54 families	N/A	A allele	Causal	(Wright et al., 2009)
Amelogenesis imperfecta-hypomaturation type IIA1	*KLK4*	Mutation: g.2142 G>A (Trp153Stop)	DNA sequencing, truncated protein	Caucasian	71 families	N/A	A allele	Causal	(Wright et al., 2011)
Atopic dermatitis	*KLK7*	Polymorphism: 3' untranslated region AACC insertion g.5784	DNA sequencing, increased mRNA expression	Caucasian	103	261	AACC insertion	Increased risk: 2.31 (1.42–3.76), p=0.0007	(Vasilopoulos et al., 2004)
Intracranial aneurysm	19q13	Microsatellite markers (linkage)	Whole-genome amplification using primer extension preamplification	Finnish	444		Linkage to region	LOD score 3.50, p=0.00006	(van der Voet et al., 2004)
Intracranial aneurysm	*KLK* region	Polymorphisms: 18 tag SNPs	Whole-genome amplification using primer extension preamplification	Finnish	368	392	rs1722561	Increased risk: p=0.003	(Weinsheimer et al., 2007)
				Finnish	368	392	rs1701946	Increased risk: p=0.002	

Table 3 (Continued)

Outcome	Gene	Polymorphism/mutation	Genotyping method, functional effect	Ethnicity	Cases	Controls	Association	Risk estimates OR (95% CI)	References
	KLK8	Polymorphisms: rs1722561, rs1701946		Finnish/ Russian (stage 2)	524	578	rs1722561	Increased risk: 1.35 (1.14–1.60), p=0.0005	
				Finnish/ Russian (stage 2)	524	578	rs1701946	Increased risk: 1.32 (1.12–1.57), p=0.0011	
Peutz-Jeghers syndrome	19q13	Microsatellite markers (linkage)	PCR and autoradiography	Caucasian/ Indian	6 families		Linkage to region	LOD score 3.80	(Mehenni et al., 1997)

Table 3 *KLK* genetic associations performed in non-malignant diseases.
CI, confidence interval; OR, odds ratio; PCR, polymerase chain reaction; RFLP, restriction fragments length polymorphism. Blood samples have been used for DNA extraction and genotyping in above studies, unless otherwise specified. *Blood and urine DNA.

rs1058205 and rs2735839) showing significant associations with prostate cancer risk and significantly associating with PSA levels in Afro-American men (Bensen et al., 2012), as detailed in Table 4.

Even though no significant polymorphisms in *KLK4* were identified relating to prostate cancer susceptibility, recent efforts by a multistage GWAS i.e., stages I and II of the Cancer Genetics Markers of Susceptibility Initiative have established gene interactions of rs2735839 (taken as a conditioning SNP) with rs1558874, which is intronic to *PRRX2* (p=4.80E-5, multiplicative OR =1.33), and rs17714461 (p=7.14E-4) located 15 kb from *KLK4* and 60 kb from the conditioning SNP, although not demonstrating LD with it (r^2 <0.001). Other notable interactions of rs2735839 were with rs17714461 (close to KLK4) and rs1558875, an intronic SNP to *PRRX2*, both SNPs being implicated in prostate cancer cell proliferation by previous studies (Ciampa et al., 2011).

A study of the *KLK12* gene locus (cases=3153 and controls=3199) established an association between rs3865443, a SNP in *KLK12* having a marginal statistically significant association with prostate cancer risk, and a requirement for further validation in a larger sample group (Supplementary Table 3).

Another recent interesting finding was that *KLK14* was inversely androgen regulated in prostate cancer cells (Lose et al., 2012). The potential of this gene in prostate cancer prognosis is promising, with the further identification of three SNPs around the *KLK14* locus being associated with prostate cancer aggression. The SNPs were rs17728459 and rs4802765, located 9 kb and 2 kb upstream of *KLK14*, respectively, and rs35287116, which encodes a p.Gln33Arg substitution in the *KLK14* signal peptide region (Lose et al., 2012) listed in Table 4.

In the process of GWAS follow-up, our study of the *Prostinogen* (*KLK15*) gene identified rs2659056 to be associated with tumor aggressiveness in a Queensland (QLD) study cohort (cases=1011 and controls=1405). This was again confirmed in replicate sets of UK GWAS stage 3 study. A highly significant association with Gleason score was observed in the combined analysis from different datasets (Table 4; Batra et al., 2011a). However, further experiments are needed to establish the functional relevance of the RORalpha transcription-factor binding alteration predicted *in silico* by the study (Batra et al., 2011a).

The intrinsic design of the GWAS means that significantly associated SNPs are seldom those that are causally linked to the phenotype, and are instead in LD with a functionally important variant. Identification of the causal variant is important for understanding of the

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
KLK2	rs198977 (C792T)	Sequenom MassARRAY				1389	1615	Yes	TT/CT: OR 1.16 (1.0–1.3); p=0.05	(Nam et al., 2009)
		Microarray				703	–	Yes, with biochemical recurrence	T allele: OR 1.58 (1.15–2.18); p<0.01	(Morote et al., 2011)
		PCR and fluorescence probe hybridization				182	–	Yes, with decreased Gleason score	p=0.04	(Kohli et al., 2010)
		PCR and sequencing (big dye terminator chemistry)				2686	1637	Yes	OR 1.08 (0.97–1.19); p=0.029	(Klein et al., 2010)
KLK2	rs2664155	Sequenom MassARRAY	G/A	TFBS		1389	1615	Yes	AG/AA: OR 1.24 (1.1–1.4); p=0.001	(Nam et al., 2009)
		Sequenom MassARRAY				1030	1327	Yes	With decreased risk; p=0.03	(Penney et al., 2011)
KLK3	rs266882	Sequenom MassARRAY (paraffin embedded seminal vesicle tissue used)				1224	–	Yes, lymph node invasion	G allele: increased lymph node invasion; p=0.02	(Cramer et al., 2008)
		Sequenom MassARRAY				1030	1327	No	Not significant with risk, clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK3	rs925013	Sequenom MassARRAY	G/A	TFBS		1224	–	Yes, with Gleason score	G allele: increased % 4 or 5 Gleason score; p=0.03	(Cramer et al., 2008)
	(G-4643A)	Sequenom MassARRAY				1030	1327	No	Not significant with risk, clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK3	rs266849	Illumina Infinium HumanHap550 array	A/G		Stage 1	1854	1894	Yes	Per allele OR 0.62 (0.55–0.69); $p=1\times10^{-16}$	(Eeles et al., 2008)
		5′-Nuclease assay (TaqMan™)			Stage 2	3268	3366	No	Not validated in stage 2	
		TaqMan™ assay				10 015	10 348	Yes	Per allele OR 0.93 (0.89–0.98); p=0.0085	(Lindstrom et al., 2011)
		5′-Endo nuclease assay (TaqMan™)			Stage 1	1854	1894	Yes	Per allele OR 0.62 (0.55–0.69); $p=1.7\times10^{-16}$	(Kote-Jarai et al., 2011a)
		5′-Endo nuclease assay (TaqMan™)			Stage 2	3650	3940	No	Not validated in stage 2	
		Illumina Golden Gate Assay			Stage 3	4901	4847	Yes	Per allele OR 0.81 (0.74–0.87); overall combined $p=1.4\times10^{-14}$	

Table 4 (Continued)

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
KLK3	rs2735839	Illumina Infinium HumanHap550 array	A/G		Stage 1	1854	1894	Yes	Per allele OR 0.56 (0.50–0.64); $p=2.4\times10^{-20}$	(Eeles et al., 2008)
		5′-Nuclease assay (Taqman™)			Stage 2	3268	3366	Yes	Per allele OR 0.83 (0.75–0.91); combined $p=1.5\times10^{-18}$	
		Sequenom MassARRAY (normal seminal vesicle tissue used)				1563	–	Yes, with aggressiveness	With less aggressive cases; $p=0.03$, not significant after multiple testing correction	(Xu et al., 2008)
								Yes, with stage	With lower stage; $p=0.03$ Not significant for age at diagnosis	
		Illumina Sentrix HumanHap550 BeadChip technology				169	805	No	Not significant for risk, aggressiveness or early onset	(Camp et al., 2009)
		Sequenom MassARRAY				1389	1615	Yes	AG/AA: OR 1.24 (1.1–1.4); $p=0.001$	(Nam et al., 2009)
		Sequenom MassARRAY (normal seminal vesicle tissue used)				5895	–	Yes, with aggressiveness	With less aggressive disease OR 1.69 (1.22–2.36); $p=0.002$	(Kader et al., 2009)
								Yes	With lower Gleason grade: $p=3.7\times10^{-7}$	
								Yes	With lower stage; $p=1.9\times10^{-4}$	
								No	Not significant for age at diagnosis	
		SNPlex Genotyping System (Applied Biosystems)				1308	1267	Yes	Per allele OR 0.84 (0.72–0.99); $p=0.04$	(Fitzgerald et al., 2009)
		Sequenom MassARRAY				454, African-American	301, African-American	Yes	Per allele OR 0.78 (0.60–1.00); $p=0.04$	(Hooker et al., 2010)
								Yes, with prostate-cancer specific death	Per A allele: HR 1.65 (1.18–2.30); $p=0.003$	
		TaqMan™ assay				9862	10 366	Yes	Per allele OR: 0.87 (0.82–0.92); $p=3.05\times10^{-6}$	(Lindstrom et al., 2011)

Table 4 (Continued)

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
								Yes, with Gleason grade	p=0.0001, not significant for stage	
		TaqMan™ assay				3522	3338	No	Per allele OR 0.93 (0.83–1.04); p=0.19	(Parikh et al., 2011)
		5′-Endo nuclease assay (TaqMan™)			Stage 1	1854	1894	Yes	Per allele OR 0.56 (0.50–0.64); $p=8.2\times10^{-20}$	(Kote-Jarai et al., 2011a)
		5′-Endo nuclease assay (TaqMan™)			Stage 2	3650	3940	Yes	Per allele OR 0.84 (0.77–0.93); combined $p=2.3\times10^{-17}$	
		Illumina Golden Gate Assay			Stage 3	4901	4847	Yes	Per allele OR 0.80 (0.73–0.88); overall combined $p=1.1\times10^{-22}$	
KLK3	rs1058205	5′-Endo nuclease assay (TaqMan™)	T/C	miRNA binding site	Stage 1	1854	1894	Yes	Per allele OR 0.59 (0.52–0.66); $p=4.7\times10^{-20}$	(Kote-Jarai et al., 2011a)
		5′-Endo nuclease assay (TaqMan™)			Stage 2	3650	3940	Yes	Per allele OR 0.85 (0.78–0.93); combined $p=1.6\times10^{-17}$	
		Illumina Golden Gate Assay			Stage 3	4901	4847	Yes	Per allele OR 0.81 (0.75–0.88); overall combined $p=2.8\times10^{-28}$	
		Sequenom MassArray				1030	1327	Yes	With decreased risk; p=0.03 Not significant with clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK3	rs17632542		T/C	Splicing (ESE or ESS), nsSNP		5325	41 417	Yes, with age at diagnosis	T allele: p=0.016	(Gudmundsson et al., 2010)
	Ile179Thr							Yes, with aggressiveness	T allele: OR 0.78; p=0.0099	
		TaqMan™ assay				3522	3338	Yes	Per allele OR 0.77 (0.67–0.89), p=0.000341	(Parikh et al., 2011)
		5′-Endo nuclease assay (TaqMan™)			Stage 1	1854	1894	Yes	Per allele 0.35 (0.30–0.42); $p=2.9\times10^{-29}$	(Kote-Jarai et al., 2011a)
		5′-Endo nuclease assay (TaqMan™)			Stage 2	3650	3940	Yes	Per allele 0.78 (0.68–0.90); combined $p=1.6\times10^{-24}$	

Table 4 (Continued)

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
		Illumina Golden Gate Assay			Stage 3	4901	4847	Yes	Per allele OR 0.65 (0.57–0.74), overall combined $p=3.9\times10^{-22}$	
KLK3	rs62113212	TaqMan™ assay	C/T			3522	3338	Yes	Per allele OR 0.79 (0.69–0.91); $p=0.00117$	(Parikh et al., 2011)
KLK3	rs62113214	TaqMan™ assay	T/G			3522	3338	Yes	Per allele OR 0.77 (0.67–0.89); $p=0.000357$	(Parikh et al., 2011)
		Sequenom MassArray				1030	1327	No	Not significant for risk, clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK3	rs2292186	Sequenom MassArray	G/A	TFBS		1030	1327	No	Not significant for risk, clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK3	rs61752561		G/A			798 Ashkenazi Jewish	–	Yes, with PrCa specific death	Per A allele HR 3.1 (1.84–5.20); $p<0.0005$	(Gallagher et al., 2010)
								Yes, with metastasis	Per A allele HR 2.16 (1.20–3.90); $p=0.011$	
KLK3	rs56397626	TaqMan™ assay	T/C			3522	3338	Yes	Per allele OR 0.80 (0.71–0.92); $p=0.000495$	(Parikh et al., 2011)
KLK3	rs11665698	TaqMan™ assay	C/A	TFBS		3522	3338	Yes	Per allele OR 0.88 (0.82–0.95); $p=0.00144$	(Parikh et al., 2011)
KLK3	rs2659124	TaqMan™ assay	T/A	TFBS		3522	3338	No	Not significant after correction for multiple testing	(Parikh et al., 2011)
KLK3	rs266878	TaqMan™ assay	C/G	TFBS		3522	3338	No	Not significant after correction for multiple testing	(Parikh et al., 2011)
		Sequenom MassArray				1030	1327	Yes	With decreased risk: $p=0.02$ Not significant with clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)

Table 4 (Continued)

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
KLK3	rs174776	TaqMan™ assay	C/T	miRNA binding site		3522	3338	No	Not significant after correction for multiple testing	(Parikh et al., 2011)
		Sequenom MassArray				1030	1327	Yes	With decreased risk: p=0.045	(Penney et al., 2011)
KLK3	rs266877	Sequenom MassArray	G/A			1030	1327	Yes	With decreased risk: p=0.02 Not significant with clinical stage, mortality or incidence of lethal disease	(Penney et al., 2011)
KLK15/*KLK3*	rs266870	Illumina Golden Gate Array (blood and buccal cell DNA)	T/T[a]		African American (AA)	1060		Yes	OR=0.8 (0.7–1.0), p=0.049, from linear mixed model	(Bensen et al., 2012)
		Illumina Golden Gate array	T/T		European American (EA)	1087		Yes	OR= 1.2 (1.0–1.5), p=0.015, from linear mixed model	(Bensen et al., 2012)
KLK3	rs1058205	Illumina Golden Gate array	T/T		AA	1060		Yes	OR= 0.8 (0.6–0.9), p=0.004, from linear mixed model	(Bensen et al., 2012)
		Illumina Golden Gate array	T/T		EA	1087		No	OR= 0.9 (0.7–1.2), p=0.485, from linear mixed model	(Bensen et al., 2012)
KLK3/*KLK2*	rs2735839	Illumina Golden Gate array	G/G		AA	1060		Yes	OR= 0.8 (0.6–0.9), p=0.005, from linear mixed model	(Bensen et al., 2012)
		Illumina Golden Gate array	T/T		EA	1087		No	OR= 0.9 (0.7–1.1), p=0.248 from linear mixed model	(Bensen et al., 2012)
KLK12	rs3865443	Sequenom MassARRAY	G/T	TFBS	Combined Australia and UK sample sets	3153	3199	Yes	TT: OR 1.28 (1.04–1.57); p=0.018	(Lose et al., 2011)
KLK12	rs3745540	Sequenom MassARRAY	C/T	TFBS		1011	1338	No	Not significant for risk or aggressiveness	(Lose et al., 2011)
KLK13	rs2736433	Sequenom MassARRAY	C/T	Splicing (ESE or ESS)		1011	1338	Yes	Decreased risk: p=0.032	(Lose et al., 2011)

Table 4 (Continued)

Gene	SNP	Genotyping method	Allele	Putative functional role*	Stage	Cases	Controls	Association?	Risk estimates (95% CI)	References
KLK13									Aggressive disease TT: 1.57 (1.00–2.48); p=0.051	
KLK13	rs2569475	Sequenom MassARRAY	C/T			1011	1338	Yes, with aggressiveness	Per allele 0.75 (0.59–0.95); p=0.0018, not replicated in UK dataset	(Lose et al., 2011)
KLK13	rs2569474	Sequenom MassARRAY	T/C			1011	1338	Yes, with aggressiveness	Per allele 0.718 (0.57–0.91); p=0.005, not replicated in UK dataset	(Lose et al., 2011)
KLK13	rs8111207	Sequenom MassARRAY	T/A			1011	1338	Yes, with aggressiveness	AA: OR 1.77 (1.07–2.91); p=0.025, not replicated in UK dataset	(Lose et al., 2011)
KLK14	rs17728459	Sequenom MassARRAY	C/T			1094	1087	Yes, decreased aggressiveness	0.33 (0.15–0.73) p=0.006, associated with Gleason score	(Lose et al., 2012)
KLK14	rs4802765	Sequenom MassARRAY	T/C			1265	1220	Yes, increased aggressiveness	1.31 (1.00–1.72) p=0.050, associated with Gleason score	(Lose et al., 2012)
KLK14	rs35287116	Sequenom MassARRAY	T/C	Substitution p.Gln33Arg in the signal peptide region		1264	1238	Yes, increased aggressiveness	1.28 (1.06–1.56) p=0.012, associated with Gleason score	(Lose et al., 2012)
KLK15	rs2659056	Illumina Infinium HumanHap550 array	A/G	TFBS	Stage 1	1854	1894	Yes	Per allele OR 1.33 (1.20–1.49); $p=1.2\times10^{-7}$	(Eeles et al., 2008)
		5'-Nuclease assay (TaqMan™)			Stage 2	3268	3366	No	Not validated by stage 2	
		Sequenom MassARRAY			Combined Australian, UK and USA study set cases	5074	–	Yes, with aggressiveness	OR 0.85 (0.77–0.93); $p=2.7\times10^{-4}$	(Batra et al., 2011a)

Table 4 Summary of *KLK* SNP association studies performed in prostate cancer post genome-wide association studies (2008 onwards).
*As predicted by 'FuncPred' from the SNPinfo web-server (http://manticore.niehs.nih.gov/snpfunc.htm). CI, confidence interval; OR, odds ratio; PCR, polymerase chain reaction; RFLP, restriction fragments length polymorphism; TFBS, transcription factor binding site. Blood samples have been used for DNA extraction and genotyping in above studies, unless otherwise specified.

molecular mechanisms underlying the pathogenesis of disease. Moreover, sincere efforts are being made to include various 'omics' levels from a systems biology viewpoint (Quigley and Balmain, 2009), epigenetic considerations and gene-environment interactions to boost casual variant determination and also in the follow-up functional analysis of isolated SNPs in the post-GWAS era (Wang et al., 2011). Consequently, additional intensive studies are required to complement GWAS to identify disease-causing alleles and their functional role in pathogenicity, such as fine-mapping and imputation studies.

Fine-mapping studies on GWAS identified *KLK* SNPs

One method that can be used to identify causal SNPs involves the performance of fine-mapping studies. These involve examining the association of all known common sequence variants in the vicinity of the GWAS-identified SNP with the disease of interest. Appropriate common sequence variants may be identified by accessing SNP databases, using sequencing data from the 1000 Genomes project, or by performing re-sequencing studies of the region of interest.

As discussed previously, Parikh et al. (2010) performed deep sequencing analysis of a 56 kb region flanking the GWAS-identified prostate cancer risk SNP rs2735839 in *KLK3*. Using these sequencing results, they then selected 24 SNPs to tag the region surrounding rs2735839 and genotyped these in five prostate cancer case-control studies from the US, France, Norway and Finland (cases=3522 and controls=3338; Parikh et al., 2011). While no strong association was observed with the original *KLK3* variant, rs2735839 (p=0.20), there was evidence of an association of three highly-correlated SNPs (rs17632542, rs62113212 and rs62113214) with prostate cancer risk (rs17632542 per allele OR 0.7, 95% CI =0.67–0.89, $p=3.41\times10^{-4}$). When stratified by disease aggressiveness, an association was only observed among less severe prostate cancer cases (Gleason score <7 and disease stage <III). The rs17632542 SNP is potentially functional, introducing a non-synonymous amino acid change from isoleucine (hydrophobic) to threonine (polar) at position 179 of the KLK3 protein. This amino acid is conserved in humans, chimpanzee and rhesus monkeys but not in other mammals or vertebrates. It is unclear whether this amino acid change has a benign or neutral functional impact and it is currently being investigated (Parikh et al., 2011).

Imputation: a new tool for fine-mapping studies

Another method that can be used to refine GWAS signals and identify causal SNPs is imputation. Genotype imputation is the process of predicting (or imputing) genotypes for known variants that are not directly assayed in a sample of individuals. These ungenotyped variants can then be tested for association with the trait. Imputation involves the comparison of study samples genotyped for a relatively large number of genetic markers (100 000–1 000 000 SNPs) to a reference panel of haplotypes derived from a number of individuals genotyped at all markers of interest (Browning, 2008). To date, the HapMap database has typically served as this reference panel, with Phase II of this project (Frazer et al., 2007) including over 3.1 million SNPs genotyped on four panels of individuals. Other reference panels, such as the 1000 Genomes project, have recently been made available.

A recent paper by Kote-Jarai et al. (2011a) undertook an imputation approach to refine the association between SNPs in the *KLK3* GWAS-identified region and prostate cancer. Using genotyping data from a two-stage GWAS using British and Australian samples (Eeles et al., 2009) and the Cancer Genetic Markers of Susceptibility study (http://www.cgems.cancer.gov/), genotypes were imputed for 197 and 312 SNPs from HapMap Phase II and the 1000 Genomes project, respectively. Interestingly, the same previously unreported SNP identified in the fine mapping study by Parikh et al., in 2011, rs17632542, was also found to be strongly associated with prostate cancer risk in this study. The association was subsequently confirmed by direct genotyping of 10 405 cases and 10 681 control individuals from the three stages of the British/Australian GWAS. This association remained strong after adjusting for the GWAS-identified SNP rs2735839 ($p=8.5\times10^{-14}$). The authors suggest rs17632542 to be the most plausible functional variant and multiple molecular dynamic simulations revealed that the threonine variant displayed superior stability in solution with likely displacement of the kallikrein loop (Kote-Jarai et al., 2011a). The functional consequences of these *in silico* findings are yet to be established.

Aside from prostate cancer, the *KLK* locus has not specifically been identified by GWAS to be associated with any other disease or trait. Coverage of genetic variation in the *KLK* locus by the genotyping chips used in these studies is quite poor, however, ranging from just 6% of the SNPs in the *KLK9* gene to 55% in the *KLK14* gene (Lose and Batra, unpublished data). Hence, more comprehensive and targeted investigations of SNPs in

the *KLK* locus and various diseases are still warranted. In addition, the majority of GWAS studies have been performed in Caucasian subjects, thus genome-wide studies of Asian and African/African-American populations may yet reveal or confirm a role for the *KLKs* in other diseases.

Association studies on high-risk variants in *KLK* genes

Rare genetic variants are considered to have a larger phenotypic effect on disease risk. As lethal variations are not favorable to propagation through natural selection, many of these have been identified using family-based studies (Marian, 2012). Rare variants were commonly overlooked in GWAS analysis due to their inability to reach statistical significance relative to other common variants. Thus there was a conceptual shift of hypothesis from common disease to common variant (CD–CV) to rare variants to common disease (RV–CD). This has been guiding next-generation sequencing measures to direct focus towards high-risk variants in order to establish complex disease associations (Marian, 2012).

Only one *KLK* gene, *KLK4*, has successfully been identified as containing a high-risk, disease-causing mutation. Amelogenesis imperfecta (AI)–hypomaturation type IIA1, as listed in Table 3, is a disorder of the teeth that results from incomplete mineralization of tooth enamel. *KLK4* was shown to be one of several factors critical for normal enamel formation (Simmer and Hu, 2002) and was therefore investigated for mutations in AI families. The initial study identified a truncating mutation, p.W153X, in one AI family that occurs at a highly conserved tryptophan residue and results in a KLK4 protein lacking the S207 residue of the catalytic triad (Hart et al., 2004). Two further AI studies carried out using DNA from blood and saliva samples have identified several more Caucasian families carrying this mutation (Wright et al., 2009, 2011), cementing *KLK4* as a high-risk gene for this condition.

Conclusions

Genetic variation in the *kallikrein* genes has been a popular focus for human disease research since their discovery. Initially, many conflicting results were observed, a phenomenon typical of candidate gene studies at the time. Since the advent of genome-wide association studies involving large numbers of patients and controls, quite promising results have arisen, particularly those implicating SNPs around *KLK2* and *KLK3* in prostate cancer risk. Although it is not likely that *KLK* SNP(s) alone will be useful in a clinical setting without the incorporation of SNPs from other loci and/or additional variables, particular *KLK* SNPs may one day form part of rapid, germline DNA-based clinical tests that may be able to stratify patients at a high risk of developing prostate cancer from those with a very low prostate cancer risk.

Acknowledgements: The current work is supported by NHMRC grant #1009458; NHMRC Principal Research Fellowship (JAC) and NHMRC Early Career Fellowship 1013693 (JB.), QLD. Government Smart State award (TOM), Australian Postgraduate Award (TOM), and the Institute of Health and Biomedical Innovation (JB, TOM).

Received May 24, 2012; accepted July 30, 2012

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7,* 248–249.

Ahn, J., Berndt, S.I., Wacholder, S., Kraft, P., Kibel, A.S., Yeager, M., Albanes, D., Giovannucci, E., Stampfer, M.J., Virtamo, J., et al. (2008). Variation in KLK genes, prostate-specific antigen and risk of prostate cancer. Nat. Genet. *40,* 1032–1034; author reply 1035–1036.

Antoniou, A.C., Beesley, J., McGuffog, L., Sinilnikova, O.M., Healey, S., Neuhausen, S.L., Ding, Y.C., Rebbeck, T.R., Weitzel, J.N., Lynch, H.T., et al., (2010). Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. Cancer Res. *70,* 9742–9754.

Azizi, M., Boutouyrie, P., Bissery, A., Agharazii, M., Verbeke, F., Stern, N., Bura-Riviere, A., Laurent, S., Alhenc-Gelas, F., and Jeunemaitre, X. (2005). Arterial and renal consequences of partial genetic deficiency in tissue kallikrein activity in humans. J. Clin. Invest. *115,* 780–787.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21,* 263–265.

Batra, J., Lose, F., O'Mara, T., Marquart, L., Stephens, C., Alexander, K., Srinivasan, S., Eeles, R.A., Easton, D.F., Al Olama, A.A., et al. (2011a). Association between prostinogen (KLK15) genetic variants and prostate cancer risk and aggressiveness in Australia and a meta-analysis of GWAS data. PLoS ONE *6,* e26527.

Batra, J., Nagle, C.M., O'Mara, T., Higgins, M., Dong, Y., Tan, O.L., Lose, F., Skeie, L.M., Srinivasan, S., Bolton, K.L., et al. (2011b). A kallikrein 15 (KLK15) single nucleotide polymorphism located close to a novel exon shows evidence of association with poor ovarian cancer survival. BMC Cancer *11,* 119.

Bensen, J.T., Xu, Z., Smith, G.J., Mohler, J.L., Fontham, E.T.H., and Taylor, J.A. (2012). Genetic polymorphism and prostate cancer aggressiveness: a case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. Prostate May 1 [Epub ahead of print].

Bharaj, B.B., Luo, L.Y., Jung, K., Stephan, C., and Diamandis, E.P. (2002). Identification of single nucleotide polymorphisms in the human kallikrein 10 (KLK10) gene and their association with prostate, breast, testicular, and ovarian cancers. Prostate *51,* 35–41.

Binnie, M.C., Alexander, F.E., Heald, C., and Habib, F.K. (2005). Polymorphic forms of prostate specific antigen and their interaction with androgen receptor trinucleotide repeats in prostate cancer. Prostate *63,* 309–315.

Biros, E., Norman, P.E., Walker, P.J., Nataatmadja, M., West, M., and Golledge, J. (2011). A single nucleotide polymorphism in exon 3 of the kallikrein 1 gene is associated with large but not small abdominal aortic aneurysm. Atherosclerosis *217,* 452–457.

Brookes, A.J. (1999). The essence of SNPs. Gene *234,* 177–186.

Browning, S.R. (2008). Missing data imputation and haplotype phase inference for genome-wide association studies. Hum. Genet. *124,* 439–450.

Camp, N.J., Farnham, J.M., Wong, J., Christensen, G.B., Thomas, A., and Cannon-Albright, L.A. (2009). Replication of the 10q11 and Xp11 prostate cancer risk variants: results from a Utah pedigree-based study. Cancer Epidemiol. Biomarkers Prev. *18,* 1290–1294.

Chiang, C.H., Chen, K.K., Chang, L.S., and Hong, C.J. (2004). The impact of polymorphism on prostate specific antigen gene on the risk, tumor volume and pathological stage of prostate cancer. J. Urol. *171,* 1529–1532.

Chiang, C.H., Hong, C.J., Chang, Y.H., Chang, L.S., and Chen, K.K. (2005). Human kallikrein-2 gene polymorphism is associated with the occurrence of prostate cancer. J. Urol. *173,* 429–432.

Ciampa, J., Yeager, M., Amundadottir, L., Jacobs, K., Kraft, P., Chung, C., Wacholder, S., Yu, K., Wheeler, W., Thun, M.J., et al. (2011). Large-scale exploration of gene–gene interactions in prostate cancer using a multistage genome-wide association study. Cancer Res. *71,* 3287–3295.

Cicek, M.S., Liu, X., Casey, G., and Witte, J.S. (2005). Role of androgen metabolism genes CYP1B1, PSA/KLK3, and CYP11alpha in prostate cancer risk and aggressiveness. Cancer Epidemiol. Biomarkers Prev. *14,* 2173–2177.

Clements, J.A. (2008). Reflections on the tissue kallikrein and kallikrein-related peptidase family – from mice to men – what have we learnt in the last two decades? Biol. Chem. *389,* 1447–1454.

Collins, F.S., Brooks, L.D., and Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. *8,* 1229–1231.

Cramer, S.D., Sun, J., Zheng, S.L., Xu, J., and Peehl, D.M. (2008). Association of prostate-specific antigen promoter genotype with clinical and histopathologic features of prostate cancer. Cancer Epidemiol. Biomarkers Prev. *17,* 2451–2457.

Cunningham, J.M., Hebbring, S.J., McDonnell, S.K., Cicek, M.S., Christensen, G.B., Wang, L., Jacobsen, S J., Cerhan, J.R., Blute, M.L., Schaid, D.J., et al. (2007). Evaluation of genetic variations in the androgen and estrogen metabolic pathways as risk factors for sporadic and familial prostate cancer. Cancer Epidemiol. Biomarkers Prev. *16,* 969–978.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. *12,* 499–510.

Devlin, B. and Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics *29,* 311–322.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nat. Genet. *40,* 316–321.

Eeles, R.A., Kote-Jarai, Z., Al Olama, A.A., Giles, G.G., Guy, M., Severi, G., Muir, K., Hopper, J.L., Henderson, B.E., Haiman, C.A., et al. (2009). Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat. Genet. *41,* 1116–1121.

Fitzgerald, L.M., Kwon, E.M., Koopmeiners, J.S., Salinas, C.A., Stanford, J.L., and Ostrander, E.A. (2009). Analysis of recently identified prostate cancer susceptibility loci in a population-based study: associations with family history and clinical features. Clin. Cancer Res. *15,* 3231–3237.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449,* 851–861.

Gallagher, D.J., Vijai, J., Cronin, A.M., Bhatia, J., Vickers, A.J., Gaudet, M.M., Fine, S., Reuter, V., Scher, H.I., Hallden, C., et al. (2010). Susceptibility loci associated with prostate cancer progression and mortality. Clin. Cancer Res. *16,* 2819–2832.

Goard, C.A., Bromberg, I.L., Elliott, M.B., and Diamandis, E.P. (2007). A consolidated catalogue and graphical annotation of dbSNP polymorphisms in the human tissue kallikrein (KLK) locus. Mol. Oncol. *1,* 303–312.

Gsur, A., Preyer, M., Haidinger, G., Zidek, T., Madersbacher, S., Schatzl, G., Marberger, M., Vutuc, C., and Micksche, M. (2002). Polymorphic CAG repeats in the androgen receptor gene, prostate-specific antigen polymorphism and prostate cancer risk. Carcinogenesis *23,* 1647–1651.

Gudmundsson, J., Besenbacher, S., Sulem, P., Gudbjartsson, D.F., Olafsson, I., Arinbjarnarson, S., Agnarsson, B.A., Benediktsdottir, K.R., Isaksson, H.J., Kostic, J.P., et al. (2010). Genetic correction of PSA values using sequence variants associated with PSA levels. Sci. Transl. Med. *2,* 62ra92.

Hart, P.S., Hart, T.C., Michalec, M.D., Ryu, O.H., Simmons, D., Hong, S., and Wright, J.T. (2004). Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta. J. Med. Genet. *41,* 545–549.

Herrala, A., Kurkela, R., Porvari, K., Isomaki, R., Henttu, P., and Vihko, P. (1997). Human prostate-specific glandular kallikrein is expressed as an active and an inactive protein. Clin. Chem. *43,* 279–284.

Hindorff, L.A., Macarthur, J., Wise, A., Junkins, H.A., Ghall, P.N., Klemm, A.K., and Manolio, T.A. (2011). A Catalog of Published Genome-wide Association Studies. (US: National Human Genome Research Institute). (http://www.genome.gov/gwastudies/).

Hooker, S., Hernandez, W., Chen, H., Robbins, C., Torres, J.B., Ahaghotu, C., Carpten, J., and Kittles, R.A. (2010). Replication of prostate cancer risk loci on 8q24, 11q13, 17q12, 19q33, and Xp11 in African Americans. Prostate *70,* 270–275.

Hua, H., Zhou, S., Liu, Y., Wang, Z., Wan, C., Li, H., Chen, C., Li, G., Zeng, C., Chen, L., et al. (2005). Relationship between the regulatory region polymorphism of human tissue kallikrein gene and essential hypertension. J. Hum. Hypertens. *19,* 715–721.

Jesser, C., Mucci, L., Farmer, D., Moon, C., Li, H., Gaziano, J.M., Stampfer, M., Ma, J., and Kantoff, P. (2008). Effects of G/A polymorphism, rs266882, in the androgen response element 1 of the PSA gene on prostate cancer risk, survival and circulating PSA levels. Br. J. Cancer *99,* 1743–1747.

Kader, A.K., Sun, J., Isaacs, S.D., Wiley, K.E., Yan, G., Kim, S.T., Fedor, H., DeMarzo, A.M., Epstein, J.I., Walsh, P.C., et al. (2009). Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. Prostate *69,* 1195–1205.

Kang, S.W., Shih, P.A., Mathew, R.O., Mahata, M., Biswas, N., Rao, F., Yan, L., Bouchard, J., Malhotra, R., Tolwani, A., et al. (2011). Renal kallikrein excretion and epigenetics in human acute kidney injury: expression, mechanisms and consequences. BMC Nephrol. *12,* 27.

Klein, R.J., Hallden, C., Cronin, A.M., Ploner, A., Wiklund, F., Bjartell, A.S., Stattin, P., Xu, J., Scardino, P.T., et al. (2010). Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer. Cancer Prev. Res. (Phila.) *3,* 611–619.

Kohli, M., Rothberg, P.G., Feng, C., Messing, E., Joseph, J., Rao, S.S., Hendershot, A., and Sahsrabudhe, D. (2010). Exploratory study of a KLK2 polymorphism as a prognostic marker in prostate cancer. Cancer Biomark. *7,* 101–108.

Kote-Jarai, Z., Easton, D.F., Stanford, J.L., Ostrander, E.A., Schleutker, J., Ingles, S.A., Schaid, D., Thibodeau, S., Dork, T., Neal, D., et al. (2008). Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. Cancer Epidemiol. Biomarkers Prev. *17,* 2052–2061.

Kote-Jarai, Z., Amin Al Olama, A., Leongamornlert, D., Tymrakiewicz, M., Saunders, E., Guy, M., Giles, G.G., Severi, G., Southey, M., Hopper, J.L., et al. (2011a). Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. Hum. Genet. *129,* 687–694.

Kote-Jarai, Z., Olama, A.A.A., Giles, G.G., Severi, G., Schleutker, J., Weischer, M., Campa, D., Riboli, E., Key, T., Gronberg, H., et al. (2011b). Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat. Genet. *43,* 785–791.

Kruglyak, L. and Nickerson, D.A. (2001). Variation is the spice of life. Nat. Genet. *27,* 234–236.

Ladiges, W., Kemp, C., Packenham, J., and Velazquez, J. (2004). Human gene variation: from SNPs to phenotypes. Mutat. Res. *545,* 131–139.

Lai, J., Kedda, M.A., Hinze, K., Smith, R.L., Yaxley, J., Spurdle, A.B., Morris, C.P., Harris, J., and Clements, J.A. (2007). PSA/KLK3 AREI promoter polymorphism alters androgen receptor binding and is associated with prostate cancer susceptibility. Carcinogenesis *28,* 1032–1039.

Lawrence, M.G., Veveris-Lowe, T.L., Whitbread, A K., Nicol, D.L., and Clements, J.A. (2007). Epithelial-mesenchymal transition in prostate cancer and the potential role of kallikrein serine proteases. Cells Tissues Organs *185,* 111–115.

Lawrence, M.G., Lai, J., and Clements, J.A. (2010). Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. Endocr. Rev. *31,* 407–446.

Lee, S.-H. and Lee, S. (2011). Genetic association study of a single nucleotide polymorphism of kallikrein-related peptidase 2 with male infertility. Clin. Exp. Reprod. Med. *38,* 6–9.

Lee, W., Yue, P., and Zhang, Z. (2009a). Analytical methods for inferring functional effects of single base pair substitutions in human cancers. Hum. Genet. *126,* 481–498.

Lee, J.Y., Park, A.K., Lee, K.M., Park, S.K., Han, S., Han, W., Noh, D.Y., Yoo, K.Y., Kim, H., Welch, R., et al. (2009b). Candidate gene approach evaluates association between innate immunity genes and breast cancer risk in Korean women. Carcinogenesis *30,* 1528–1531.

Lee-Chen, G.J., Liu, K.P., Lai, Y.C., Juang, H.S., Huang, S.Y., and Lin, C.Y. (2004). Significance of the tissue kallikrein promoter and transforming growth factor-beta1 polymorphisms with renal progression in children with vesicoureteral reflux. Kidney Int. *65,* 1467–1472.

Lindstrom, S., Schumacher, F., Siddiq, A., Travis, R.C., Campa, D., Berndt, S.I., Diver, W.R., Severi, G., Allen, N., Andriole, G., et al. (2011). Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers – results from BPC3. PLoS ONE *6,* e17142.

Lose, F., Batra, J., O'Mara, T., Fahey, P., Marquart, L., Eeles, R.A., Easton, D.F., Al Olama, A.A., Kote-Jarai, Z., Guy, M., et al. (2011). Common variation in kallikrein genes KLK5, KLK6, KLK12, and KLK13 and risk of prostate cancer and tumor aggressiveness. Urol. Oncol. Jul 7. [Epub ahead of print]

Lose, F., Lawrence M.G., Srinivasan, S., O'Mara, T., Marquart, L., Chambers, S., Gardiner R.A., Aitken J.F., Spurdle, A.B., Batra, J. et al. (2012). The kallikrein 14 gene is down-regulated by androgen receptor signalling and harbours genetic variation that is associated with prostate tumour aggressiveness. Biol. Chem. *393*, 403.

Mardis, E.R. (2008a). The impact of next-generation sequencing technology on genetics. Trends Genet. *24*, 133–141.

Mardis, E.R. (2008b). Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. *9,* 387–402.

Marian, A.J. (2012). Molecular genetic studies of complex phenotypes. Transl. Res. *159*, 64–79.

Medeiros, R., Morais, A., Vasconcelos, A., Costa, S., Pinto, D., Oliveira, J., Carvalho, R., and Lopes, C. (2002). Linkage between polymorphisms in the prostate specific antigen ARE1 gene region, prostate cancer risk, and circulating tumor cells. Prostate *53,* 88–94.

Mehenni, H., Blouin, J.L., Radhakrishna, U., Bhardwaj, S.S., Bhardwaj, K., Dixit, V.B., Richards, K.F., Bermejo-Fenoll, A., Leal, A.S., Raval, R.C., et al. (1997). Peutz-Jeghers syndrome: confirmation of linkage to chromosome 19p13.3 and identification of a potential second locus, on 19q13.4. Am. J. Hum. Genet. *61,* 1327–1334.

Metzker, M.L. (2010). Sequencing technologies – the next generation. Nat. Rev. Genet. *11,* 31–46.

Mittal, R.D., Mishra, D.K., Thangaraj, K., Singh, R., and Mandhani, A. (2007). Is there an inter-relationship between prostate specific antigen, kallikrein-2 and androgen receptor gene

polymorphisms with risk of prostate cancer in north Indian population? Steroids *72,* 335–341.

Mononen, N., Seppala, E.H., Duggal, P., Autio, V., Ikonen, T., Ellonen, P., Saharinen, J., Saarela, J., Vihinen, M., Tammela, T.L., et al. (2006). Profiling genetic variation along the androgen biosynthesis and metabolism pathways implicates several single nucleotide polymorphisms and their combinations as prostate cancer risk factors. Cancer Res. *66,* 743–747.

Morote, J., Del Amo, J., Borque, A., Ars, E., Hernandez, C., Herranz, F., Arruza, A., Llarena, R., Planas, J., Viso, M.J., et al. (2011). Improved prediction of biochemical recurrence after radical prostatectomy by genetic polymorphisms. J. Urol. *184,* 506–511.

Nam, R.K., Zhang, W.W., Trachtenberg, J., Diamandis, E., Toi, A., Emami, M., Ho, M., Sweet, J., Evans, A., Jewett, M.A., et al. (2003). Single nucleotide polymorphism of the human kallikrein-2 gene highly correlates with serum human kallikrein-2 levels and in combination enhances prostate cancer detection. J. Clin. Oncol. *21,* 2312–2319.

Nam, R.K., Zhang, W.W., Jewett, M.A., Trachtenberg, J., Klotz, L.H., Emami, M., Sugar, L., Sweet, J., Toi, A., and Narod, S.A. (2005). The use of genetic markers to determine risk for prostate cancer at prostate biopsy. Clin. Cancer Res. *11,* 8391–8397.

Nam, R.K., Zhang, W.W., Klotz, L.H., Trachtenberg, J., Jewett, M.A., Sweet, J., Toi, A., Teahan, S., Venkateswaran, V., Sugar, L., et al. (2006). Variants of the hK2 protein gene (KLK2) are associated with serum hK2 levels and predict the presence of prostate cancer at biopsy. Clin. Cancer Res. *12,* 6452–6458.

Nam, R.K., Zhang, W.W., Trachtenberg, J., Seth, A., Klotz, L.H., Stanimirovic, A., Punnen, S., Venkateswaran, V., Toi, A., Loblaw, D.A., et al. (2009). Utility of incorporating genetic variants for the early detection of prostate cancer. Clin. Cancer Res. *15,* 1787–1793.

O'Mara, T.A., Nagle, C.M., Batra, J., Kedda, M.A., Clements, J.A., and Spurdle, A.B. (2011). Kallikrein-related peptidase 3 (KLK3/PSA) single nucleotide polymorphisms and ovarian cancer survival. Twin Res. Hum. Genet. *14,* 323–327.

Olivier, M. (2003). A haplotype map of the human genome. Physiol. Genomics *13,* 3–9.

Pal, P., Xi, H., Sun, G., Kaushal, R., Meeks, J.J., Thaxton, C.S., Guha, S., Jin, C.H., Suarez, B.K., Catalona, W.J., et al. (2007). Tagging SNPs in the kallikrein genes 3 and 2 on 19q13 and their associations with prostate cancer in men of European origin. Hum. Genet. *122,* 251–259.

Paliouras, M. and Diamandis, E.P. (2006). The kallikrein world: an update on the human tissue kallikreins. Biol. Chem. *387,* 643–652.

Parikh, H., Deng, Z., Yeager, M., Boland, J., Matthews, C., Jia, J., Collins, I., White, A., Burdett, L., Hutchinson, A., et al. (2010). A comprehensive resequence analysis of the KLK15-KLK3-KLK2 locus on chromosome 19q13.33. Hum. Genet. *127,* 91–99.

Parikh, H., Wang, Z., Pettigrew, K.A., Jia, J., Daugherty, S., Yeager, M., Jacobs, K.B., Hutchinson, A., Burdett, L., Cullen, M., et al. (2011). Fine mapping the KLK3 locus on chromosome 19q13.33 associated with prostate cancer susceptibility and PSA levels. Hum. Genet. *129,* 675–685.

Pavlopoulou, A., Pampalakis, G., Michalopoulos, I., and Sotiropoulou, G. (2010). Evolutionary history of tissue kallikreins. PLoS ONE *5,* e13781.

Pelletier, C. and Weidhaas, J.B. (2010). MicroRNA binding site polymorphisms as biomarkers of cancer risk. Exp. Rev. Mol. Diagn. *10*, 817–829.

Penney, K.L., Schumacher, F.R., Kraft, P., Mucci, L.A., Sesso, H.D., Ma, J., Niu, Y., Cheong, J.K., Hunter, D.J., Stampfer, M.J., et al. (2011). Association of KLK3 (PSA) genetic variants with prostate cancer risk and PSA levels. Carcinogenesis *32,* 853–859.

Pennisi, E. (2010). Genomics. 1000 Genomes Project gives new map of genetic diversity. Science *330,* 574–575.

Pestell, R.G. (2008). Prostate cancer: signaling networks, genetics, and new treatment strategies. Curr. Clin. Oncol. *359,* 772.

Pharoah, P.D., Antoniou, A.C., Easton, D.F., and Ponder, B.A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. N. Engl. J. Med. *358,* 2796–2803.

Prensner, J.R., Rubin, M.A., Wei, J.T., and Chinnaiyan, A.M. (2012). Beyond PSA: the next generation of prostate cancer biomarkers. Sci. Transl. Med. *4,* 127–123.

Quigley, D. and Balmain, A. (2009). Systems genetics analysis of cancer susceptibility: from mouse models to humans. Nat. Rev. Genet. *10,* 651–657.

Rao, A., Chang, B.L., Hawkins, G., Hu, J.J., Rosser, C.J., Hall, M.C., Meyers, D.A., Xu, J., and Cramer, S.D. (2003). Analysis of G/A polymorphism in the androgen response element I of the PSA gene and its interactions with the androgen receptor polymorphisms. Urology *61,* 864–869.

Salinas, C.A., Austin, M.A., Ostrander, E.O., and Stanford, J.L. (2005). Polymorphisms in the androgen receptor and the prostate-specific antigen genes and prostate cancer risk. Prostate *65,* 58–65.

Savage, S.A. (2008). Cancer genetic association studies in the genome-wide age. Per. Med. *5,* 589–597.

Schatzl, G., Marberger, M., Remzi, M., Grosser, P., Unterlechner, J., Haidinger, G., Zidek, T., Preyer, M., Micksche, M., and Gsur, A. (2005). Polymorphism in ARE-I region of prostate-specific antigen gene associated with low serum testosterone level and high-grade prostate cancer. Urology *65,* 1141–1145.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. Nat. Methods *5,* 16–18.

Severi, G., Hayes, V.M., Neufing, P., Padilla, E.J., Tilley, W.D., Eggleton, S.A., Morris, H.A., English, D.R., Southey, M.C., Hopper, J.L., et al. (2006). Variants in the prostate-specific antigen (PSA) gene and prostate cancer risk, survival, and circulating PSA. Cancer Epidemiol. Biomark. Prev. *15,* 1142–1147.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29,* 308–311.

Shinmura, K., Tao, H., Yamada, H., Kataoka, H., Sanjar, R., Wang, J., Yoshimura, K., and Sugimura, H. (2004). Splice-site genetic polymorphism of the human kallikrein 12 (KLK12) gene correlates with no substantial expression of KLK12 protein having serine protease activity. Hum. Mutat. *24,* 273–274.

Simmer, J.P. and Hu, J.C. (2002). Expression, structure, and function of enamel proteinases. Connect. Tissue Res. *43,* 441–449.

Siva, N. (2008). 1000 Genomes project. Nat. Biotechnol. *26,* 256.

Slim, R., Torremocha, F., Moreau, T., Pizard, A., Hunt, S.C., Vuagnat, A., Williams, G.H., Gauthier, F., Jeunemaitre, X., and Alhenc-Gelas, F. (2002). Loss-of-function polymorphism of the human kallikrein gene with reduced urinary kallikrein activity. J. Am. Soc. Nephrol. *13,* 968–976.

Song, Q., Chao, J., and Chao, L. (1997). DNA polymorphisms in the 5'-flanking region of the human tissue kallikrein gene. Hum. Genet. *99,* 727–734.

Tan Olivia, L., Whitbread Astrid, K., Clements Judith, A., and Dong, Y. (2006). Kallikrein-related peptidase (KLK) family mRNA variants

and protein isoforms in hormone-related cancers: do they have a function? Biol. Chem. *387*, 697.

The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.

Thomas, D.C., Haile, R.W., and Duggan, D. (2005). Recent developments in genomewide association scans: a workshop summary and review. Am. J. Hum. Genet. *77,* 337–345.

van der Voet, M., Olson, J.M., Kuivaniemi, H., Dudek, D.M., Skunca, M., Ronkainen, A., Niemela, M., Jaaskelainen, J., Hernesniemi, J., Helin, K., et al. (2004). Intracranial aneurysms in Finnish families: confirmation of linkage and refinement of the interval to chromosome 19q13.3. Am. J. Hum. Genet. *74,* 564–571.

Vasilopoulos, Y., Cork, M.J., Murphy, R., Williams, H.C., Robinson, D.A., Duff, G.W., Ward, S.J., and Tazi-Ahnini, R. (2004). Genetic association between an AACC insertion in the 3'UTR of the stratum corneum chymotryptic enzyme gene and atopic dermatitis. J. Invest. Dermatol. *123,* 62–66.

Vasilopoulos, Y., Sharaf, N., di Giovine, F., Simon, M., Cork, M.J., Duff, G.W., and Tazi-Ahnini, R. (2011). The 3'-UTR AACCins5874 in the stratum corneum chymotryptic enzyme gene (SCCE/KLK7), associated with atopic dermatitis; causes an increased mRNA expression without altering its stability. J. Dermatol. Sci. *61,* 131–133.

Wang, L.Z., Sato, K., Tsuchiya, N., Yu, J.G., Ohyama, C., Satoh, S., Habuchi, T., Ogawa, O., and Kato, T. (2003). Polymorphisms in prostate-specific antigen (PSA) gene, risk of prostate cancer, and serum PSA levels in Japanese population. Cancer Lett. *202,* 53–59.

Wang, X., Prins, B., Sõber, S., Laan, M., and Snieder, H. (2011). Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension. Curr. Hypertens. Rep. *13,* 442–451.

Weinsheimer, S., Goddard, K.A., Parrado, A.R., Lu, Q., Sinha, M., Lebedeva, E.R., Ronkainen, A., Niemela, M., Khusnutdinova, E.K., Khusainova, R.I., et al. (2007). Association of kallikrein gene polymorphisms with intracranial aneurysms. Stroke *38,* 2670–2676.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447,* 661–678.

Wright, J.T., Hart, T.C., Hart, P.S., Simmons, D., Suggs, C., Daley, B., Simmer, J., Hu, J., Bartlett, J.D., Li, Y., et al. (2009). Human and mouse enamel phenotypes resulting from mutation or altered expression of AMEL, ENAM, MMP20 and KLK4. Cells Tissues Organs *189,* 224–229.

Wright, J.T., Torain, M., Long, K., Seow, K., Crawford, P., Aldred, M.J., Hart, P.S., and Hart, T.C. (2011). Amelogenesis imperfecta: genotype-phenotype studies in 71 families. Cells Tissues Organs *194,* 279–283.

Xu, J., Meyers, D.A., Sterling, D.A., Zheng, S.L., Catalona, W.J., Cramer, S.D., Bleecker, E.R., and Ohar, J. (2002). Association studies of serum prostate-specific antigen levels and the genetic polymorphisms at the androgen receptor and prostate-specific antigen genes. Cancer Epidemiol. Biomarkers Prev. *11,* 664–669.

Xu, J., Isaacs, S.D., Sun, J., Li, G., Wiley, K.E., Zhu, Y., Hsu, F.C., Wiklund, F., Turner, A.R., Adams, T.S., et al. (2008). Association of prostate cancer risk variants with clinicopathologic characteristics of the disease. Clin. Cancer Res. *14,* 5819–5824.

Xue, W., Irvine, R.A., Yu, M.C., Ross, R.K., Coetzee, G.A., and Ingles, S.A. (2000). Susceptibility to prostate cancer: interaction between genotypes at the androgen receptor and prostate-specific antigen loci. Cancer Res. *60,* 839–841.

Xue, W.M., Coetzee, G.A., Ross, R.K., Irvine, R., Kolonel, L., Henderson, B.E., and Ingles, S.A. (2001). Genetic determinants of serum prostate-specific antigen levels in healthy men from a multiethnic cohort. Cancer Epidemiol. Biomarkers Prev. *10,* 575–579.

Yang, Q., Nakamura, M., Nakamura, Y., Yoshimura, G., Suzuma, T., Umemura, T., Tamaki, T., Mori, I., Sakurai, T., and Kakudo, K. (2002). Correlation of prostate-specific antigen promoter polymorphisms with clinicopathological characteristics in breast cancer. Anticancer Res. *22,* 1825–1828.

Yu, H., Song, Q., Freedman, B.I., Chao, J., Chao, L., Rich, S.S., and Bowden, D.W. (2002). Association of the tissue kallikrein gene promoter with ESRD and hypertension. Kidney Int. *61,* 1030–1039.

Yue, P., Melamud, E., and Moult, J. (2006). SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics *7,* 166.

Zeng, Y., Zhang, L., Hu, Z.Y., Yang, Q.D., Liu, B.Q., Ma, M.M., Liu, Y.H., Xu, H.W., Xia, J., Du, X.P., et al. (2010). Association of single nucleotide polymorphisms of kallikrein 1 gene with cerebral hemorrhage in Changsha Han Chinese. Zhonghua Yi Xue Yi Chuan Xue Za Zhi *27,* 316–319.

Zhao, W., Wang, L., Lu, X., Yang, W., Huang, J., Chen, S., and Gu, D. (2007). A coding polymorphism of the kallikrein 1 gene is associated with essential hypertension: a tagging SNP-based association study in a Chinese Han population. J. Hypertens. *25,* 1821–1827.

Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H.O., Hsu, F.C., Zhu, Y., Balter, K., et al. (2008). Cumulative association of five genetic variants with prostate cancer. N. Engl. J. Med. *358,* 910–919.

Dr Jyotsna Batra is an NHMRC Peter Doherty Fellow at Australian Prostate Cancer Research Centre-Queensland, Institute of Health and Biomedical Innovation, QUT. She has studied Biochemistry towards a Master's degree at the Kurukshetra University, India in 2002. Jyotsna finished her PhD on Human genetics in 2008 from the Institute of Genomics and Integrative Biology, India, working on the complexity of the heredity allergic disorders. The focus of her current postdoctoral fellowship is to identify the molecular consequences of prostate cancer risk-associated single nucleotide polymorphisms (SNPs) already identified by GWAS and/or by fine mapping of the *KLK3* gene, and to investigate novel Kallikrein gene SNPs extracted from Nextgen sequencing for an association with prostate cancer risk and/or prognostic features.

Tracy O'Mara graduated with a Bachelor of Applied Science (Life Sciences) with first class honours from the Queensland University of Technology (QUT) in 2005. From 2006–2008 she worked as a Research Assistant for Dr Mary-Anne Kedda from the School of Public Health and Institute of Health and Biomedical Innovation, QUT. Tracy began her PhD investigating the role of single nucleotide polymorphisms on endometrial cancer risk and aggressiveness in 2008 under the supervision of A/Prof Amanda Spurdle from the Queensland Institute of Medical Research and Prof Judith Clements from the Institute of Biomedical Innovation, QUT.

Radhika Patnala completed her B.Tech Biotechnology from GITAM University, India in 2009 and graduated with Masters of Biotechnology (Honours) from the Australian National University, Canberra in 2011, under supervision of Dr Danny Rangasamy, John Curtin School of Medical Research. She then worked as a part time Research Assistant at Queensland University of Technology with Dr Jyotsna Batra. She will be pursuing doctoral research in the field of Neuroscience from August 2012 under A/Prof S Thameem Dheen at the Yong Loo Lin School of Medicine, National University of Singapore.

Felicity Lose completed her PhD from Queensland Institute of Medical Research in 2007 in cancer genetics. Since then, she have been involved in prostate cancer research, firstly as Project Co-ordinator for two large control recruitment projects (over 1200 controls recruited total), and now involved in genotyping and analysis of our large prostate cancer case and male control sample sets for the Kallikrein genes.

BAppSc, RMIT, 1982; MAppSc, RMIT, 1983; PhD, Monash University, 1989.

Professor Judith Clements is an NHMRC Principal Research Fellow, Scientific Director of the Australian Prostate Cancer Research Centre-Queensland and leads the Cancer Program at the Institute of Health and Biomedical Innovation, Queensland University of Technology. Her research focuses on mechanistic and translational studies of the role of PSA and the related kallikrein serine peptidases in prostate cancer.