Evolution, Medicine, and Public Health [2021] pp. 311–321
doi:10.1093/emph/eoab026
Advance access date 14 September 2021

ORIGINAL RESEARCH ARTICLE

EVOLUTION, MEDICINE, & PUBLIC HEALTH

Evolutionary selection of alleles in the melanophilin gene that impacts on prostate organ function and cancer risk

Luca Ermini,[1,†] Jeffrey C. Francis,[2,†] Gabriel S. Rosa,[2] Alexandra J. Rose,[2] Jian Ning,[2,3] Mel Greaves[1,*,‡] and Amanda Swain [2,3,*,‡]

[1]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK; [2]Division of Cancer Biology, The Institute of Cancer Research, London, UK and [3]Tumour Profiling Unit, The Institute of Cancer Research, London, UK
*Corresponding author. Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. Tel: +4420 7153 5355; E-mail: Amanda.swain@icr.ac.uk (A.S.); Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, UK. Tel: + 4420 8722 4073; E-mail: mel.greaves@icr.ac.uk (M.G.)
[†]Joint first authorship.
[‡]Joint senior authorship.
Received 09 March 2021; revised version accepted 03 September 2021;

ABSTRACT

Background and objectives: Several hundred inherited genetic variants or SNPs that alter the risk of cancer have been identified through genome-wide association studies. In populations of European ancestry, these variants are mostly present at relatively high frequencies. To gain insight into evolutionary origins, we screened a series of genes and SNPs linked to breast or prostate cancer for signatures of historical positive selection.

Methodology: We took advantage of the availability of the 1000 genome data and we performed genomic scans for positive selection in five different Caucasian populations as well as one African reference population. We then used prostate organoid cultures to provide a possible functional explanation for the interplay between the action of evolutionary forces and the disease risk association.

Results: Variants in only one gene showed genomic signatures of positive, evolutionary selection within Caucasian populations melanophilin (*MLPH*). Functional depletion of *MLPH* in prostate organoids, by CRISPR/Cas9 mutation, impacted lineage commitment of progenitor cells promoting luminal versus basal cell differentiation and on resistance to androgen deprivation.

Conclusions and implications: The *MLPH* variants influencing prostate cancer risk may have been historically selected for their adaptive benefit on skin pigmentation but *MLPH* is highly expressed in the prostate and the derivative, positively selected, alleles decrease the risk of prostate cancer. Our study suggests a potential functional mechanism via which *MLPH* and its genetic variants could influence risk of prostate cancer, as a serendipitous consequence of prior evolutionary benefits to another tissue.

Lay Summary: We screened a limited series of genomic variants associated with breast and prostate cancer risk for signatures of historical positive selection. Variants within the melanophilin (*MLPH*) gene fell into this category. Depletion of *MLPH* in prostate organoid cultures, suggested a potential functional mechanism for impacting on cancer risk, as a serendipitous consequence of prior evolutionary benefits to another tissue.

KEYWORDS: human evolution; cancer susceptibility; organoids; prostate; MLPH

BACKGROUND AND OBJECTIVES

Several hundred inherited gene variants or SNPs impacting on cancer risk have been identified via genome-wide association studies (GWAS). Their individual impact on cancer risk is very modest but they are present at relatively high frequencies in European descent (risk allele frequency >5%; odds ratio < 1.5) [1]. Although the causal variants are in many cases unidentified, the functions they regulate suggest that some contribute by being active in cancer cells and perhaps epistatic with acquired mutations [2]. Approximately one-third of those cancer-associated variants have been shown to be pleiotropic for multiple cancers [1]. An example is given by telomeres-related loci at the TERT-CLPTM1L region associated with lung, bladder, prostate and cervical cancer risk [3]. Some important pleiotropic variants include those within *MYC*, *TERT* and *HNF1B*, linked to different cancer types [4].

Deaths from cancer are mostly post-reproductive and are therefore provided, at best, with only weak evolutionary selective pressure [5]. Alleles with a small impact on risk of cancer, unchecked by natural selection, could therefore increase, over time by drift, founder effects or antagonistic pleiotropy in which a trait selected for a fitness benefit, carries a post-reproductive trade off via another trait [6]. Along similar lines, a gene variant might be positively selected in one historical, environmental context but imparts a risk, including cancer, in another contemporary context. This represents an evolutionary mismatch [7] and might be applied to alleles selected 5000–10 000 years ago for skin depigmentation, associated with risk of UVB exposure and linked to skin cancer [8].

Aside from the gene variants linked to skin cancer risk, there are not many examples to date of inherited variants that are associated with cancer risk and detected SNPs that have been positively selected. The TP53 binding domain of the *Kit* ligand regulatory domain, associated with testicular cancer, was reported under positive selection [9] but this finding has been disputed [10]. Strong signature of natural selection was also reported at the *FHIT* locus that predisposes to prostate cancer [11].

Gene variants in the hormonal signalling pathways or other critical functions of breast and prostate organs might be candidates for positive selection and impact on cancer risk; some prior evidence supports this notion. African black men and Afro-American men show an increased risk of prostate cancer and have a higher frequency of androgen receptor duplications associated with adverse cancer progression [12]. An epidemiological study [13] suggested that variants within *BRCA1* may have been positively selected for a fecundity benefit. Specific *BRCA1* mutations are present at relatively high frequencies in certain populations but this has been plausibly ascribed to historical founder effects rather than selection [14].

In the light of these uncertainties, we carried out a proof of principle study to identify evolutionary positive selection in any cancer GWAS SNPs. We further sought to uncover a functional role for any such gene with cancer-associated SNPs that might explain its impact on cancer risk and its evolutionary adaptive significance.

We selected a total of 36 SNPs that were significantly linked to prostate cancer that had been fine mapped and functionally annotated [15]. Of these, 25 were SNPs within known genes and a further 11 were intergenic. Additionally, we selected a further eight genes with known critical functions in breast/prostate tissues including *BRCA1/2* and oestrogen receptors genes (*ESR1/2*). One potential candidate for selection, the androgen receptor, could not be evaluated because analysis was compromised by its single copy X chromosome status. These genes and SNPs were interrogated using different, standard summary statistics able to detect genomic signals of positive selection in humans within an estimated window of historical, evolutionary time [16]. Our objective was to seek to identify and functionally validate individual candidates that had genomic signals of positive selection, and not to determine the overall prevalence of positive selection of cancer GWAS alleles. This would require a much larger, re-iterated study.

We found none of the SNPs or gene regions evaluated had significant evidence of positive selection with two exceptions. One was in the gene *PPP1R14A* associated with increased prostate cancer risk but selected within an African population and we did not follow it up functionally. The other positively selected gene encoded *MLPH*, the derivative allele in Caucasian populations reducing prostate cancer risk [15]. This gene has a recognized function in skin pigmentation which raises the puzzle of why it might impact on cancer within the prostate. We therefore coupled our genomic analysis with an exploration of *MLPH* function in the prostate using a mouse organoid system. The results suggest a plausible functional link between the selected GWAS allele and reduced prostate cancer risk and adds to our understanding of the multiple evolutionary and genetic influences on cancer.

METHODOLOGY

Genomic data

Phased genomic data from five different European ancestry populations (Utah residents of Northern and Western European ancestry, CEU; Finnish, FIN; British, GBR; Iberian Population, IBS; Tuscans, TSI) and one African population (Yoruba, YRI, used as reference) were obtained from the 1000 Genome Project Phase 3 [17]. A total of 599 individuals were analysed (Supplementary Table S1). Thirty-six SNPs and eight genes or gene clusters were scanned for positive selection (Supplementary Tables S2 and S3).

Selection analysis

The integrated Haplotype Score (iHS) and the Cross Population Extended Haplotype Homozygosity (XP-EHH) were implemented according to previous methods [18, 19] and carried out with selscan [20]. Clustering analysis was implemented to conservatively validate significant results. Linear mixed models were used to test selection strength on *MLPH* at different latitudes. Extended haplotype homozygosity (EHH) analysis and haplotype-bifurcation diagrams (HBD), introduced by Sabeti *et al.* [21], were used to evaluate which allele carries the haplotype homozygosity and thus possibly being under positive selection. Tajima's D [22] and Fay and Wu's [23] H statistics were computed using the package PopGenome [24]. Statistical analyses were performed using R v3.6.1 [25]. Exhaustive description is provided in Supplementary Data.

MLPH expression analysis

MLPH expression was assessed by analysing RNA expression data from the Human Protein Atlas [26] and Gent2 databases [27]. A total of 12 655 and 62 978 RNA expression data were separately analysed from the first and latter database (Supplementary Table S4). Details are provided in Supplementary Data.

Prostate organoid culture

Mouse prostate tissue isolation and organoid growth were carried out as described by Drost *et al.* [28] and detailed in Supplementary Data.

CRISPR/Cas9 mutation of *Mlph*

To generate *Mlph* mutant cells, mouse prostate organoids were dissociated into single cells and transduced with lentiCRISPRv2 [29] containing *Mlph* sgRNA. Methods are provided in detail in Supplementary Data.

Organoid functional studies

Materials and Methods describing the molecular and functional analysis including *Mlph* mutant organoids genotyping, western blots, co-immunoprecipitation, immunohistochemistry (IHC), *MLPH* gene expression in prostate organoids and organoid quantification are detailed in Supplementary Data.

RESULTS

We performed genomic scans for selection in five different Caucasian populations as well as one African reference population. We combined different standard summary statistics able to detect genomic signals of positive natural selection in an estimated window of evolutionary time [16]. We first used two complementary statistics grounded on long haplotype detection and able to identify recent episodes of selective sweeps (age <30 000 years): iHS [18] and XP-EHH [19]. We then employed two different statistics grounded on the frequency expectations of the variant under neutrality: Tajima's D [22] (age <250 000 years) and Fay & Wu's H [23] (age <80 000 years).

To assess the reliability of our statistical tests, we used the *LCT/MCM6* region as control. We first computed the standardized absolute |iHS| score, (hereafter as iHS) and the standardized XP-EHH statistics for every SNP in the *LCT/MCM6* region and within each single population. Results (Supplementary Fig. S1) showed clusters of significant iHS and XP-EHH values consistent with scientific literature advocating positive selection at *LCT/MCM6* region [30]. The SNPs rs4988235 and rs182549, associated with the lactase persistence phenotype [30], as expected, show high and significant iHS scores in CEU, GBR, FIN and IBS, although in the latter dropping significance after correcting for multiple testing (Supplementary Table S5). In contrast, the distribution of iHS values in the Tuscan population does not support any selective sweep even for rs4988235 and rs182549. These two last SNPs are missing in the reference YRI dataset and therefore not analysed with XP-EHH. The graphical assessment approach, supported by EHH and HBD, suggest the derived alleles carrying the haplotype homozygosity and thus being the target of selection (Supplementary Fig. S2A) as previously demonstrated [31]. The two neutrality tests, Tajima's D and Fay and Wu's H do not provide support for positive selection [31] (Supplementary Fig. S2B) which agree with previous claims on positive selection occurring within the past 5000–10 000 years [31]. Past reports showed no evidence of selection in CEU using the same neutrality tests [32].

Selection in Europeans for the *MLPH* locus involved in prostate cancer susceptibility

The overall results are shown in Supplementary Tables S6–S8. For the majority of these genetic markers, there was no evidence for positive selection.

Only *MLPH* showed strong evidence for positive selection in European populations. One other gene, *PPP1R14A*, had strong

evidence for positive selection (rs12610267, Supplementary Fig. S3), but this was in the Yoruba population and was not selected for further study. We additionally found weaker evidence for positive selection in three further SNPs (rs2596546, rs10875943, rs10713532; Supplementary Fig. S4).

The *MLPH* gene encodes for MLPH protein, a Rab effector protein involved in melanosome transport [33]. The SNP rs11891348 (G/T ancestral/derived allele), within this gene has been associated with increasing risk of prostate cancer for those G allele carriers [15] and identified as a credible associated variant [34]. Another common SNP, rs11891426 (about 6 Kb downstream from rs11891348; G/T, ancestral/derived allele), was also associated with *MLPH* and prostate cancer risk. Carriers of the derived rs11891426 T allele showed a significantly higher MLPH expression in prostate tissues which was associated with a favourable risk profile. This suggests a causal relationship between prostate cancer development and modulation of *MLPH* expression [35].

Results based on iHS statistics showed significant rs11891348 and rs11891426 scores in all populations examined, supported by clustering analysis of the whole tested region (iHS $>$ 2.58, corrected *P* value $<$ 0.038; proportion test *P* value: 0.0013 for all) as shown in Fig. 1. In addition, results on the whole *MLPH* gene and surrounding region supports evidence for positive selection acting on the whole gene for all Europeans (Supplementary Fig. S5A).

The EHH of the derived alleles in both SNPs decays slower than the ancestral one, although showing a moderate decay (EHH distributions *P* value $<$ 0.00001). The result is supported by HBD where a long-range haplotype carrying the derived allele is present with a very thick line, suggesting high frequency. Results suggest a long-range haplotype carrying the derived and protective T allele (Fig. 1 for rs11891348; Supplementary Fig. S5B for rs11891426). Linkage Disequilibrium analysis between rs11891348 and rs11891426 indicates a significant correlation between the ancestral G alleles and the derived T alleles of both SNPs (D': 0.84; r^2: 0.50; *P* value $<$ 0.0001; Supplementary Fig. S6). The XP-EHH test supported the results obtained by iHS for all populations. Results suggested rs11891348 and rs11891426 as candidate SNPs for positive selection (XP-EHH $>$ 4.14, corrected *P* value $<$ 0.000454; proportion test *P* value: 0.0013 for all; Fig. 1). XP-EHH test performed on the whole *MLPH* suggests positive selection acting on the whole gene in all populations corroborating results obtained by iHS (Supplementary Fig. S5A). Neutrality tests for *MLPH* showed no significant results (Supplementary Fig. S5C).

A North-South latitudinal cline for selection strength

We investigated the intensity of selection acting on *MLPH* in different populations at different latitudes. Results depicted a north-south latitudinal cline for selection strength, with southern European populations being the ones most targeted by the action of positive selection (Fig. 2). The same latitudinal cline is acting on both rs11891426 and rs11891348. Furthermore, we clustered together CEU, FIN and GBR as north Europeans and IBS and TSI as south Europeans and we tested for selection strength between those two latitudinal extents. Results show significant differences between the north and south latitudes for both iHS (*P* value: 0.006644) and XP-EHH (*P* value: 0.00184) distributions.

MLPH is expressed at high levels in prostatic tissue

To further investigate the role of *MLPH*, we queried *MLPH* RNA expression datasets available within the Human Protein Atlas [26] (GTEx dataset) and Gent2 web databases [27]. Despite *MLPH* being involved in melanosome transport, the human tissue that expresses the highest levels of this gene is, on average, the prostate followed by salivary glands and stomach, while *MLPH* is expressed at a much lower level in the skin (Supplementary Fig. S7, top). A similar trend is observed in cancer tissues, where prostate cancer expresses the highest levels of *MLPH*, followed by melanoma and pancreatic cancer (Supplementary Fig. S7, bottom). We found no significant differences between the expression of *MLPH* in healthy prostate tissue and prostatic cancer (*P* value: 0.1425; Fig. 3A). A comparison between the expression of *MLPH* between two healthy tissues, skin and prostate showed a significantly higher level of expression in the latter (*P* value $< 2.2 \times 10^{-16}$; Fig. 3B). When we compared the expression of *MLPH* in skin comparing sun exposed versus sun protected skin, we found a significantly higher level of expression in the former (*P* value: 0.007774; Fig. 3C).

Functional impact of *MLPH* loss in prostate tissue

Protein expression analysis in benign and prostate cancer samples showed that tissues that contained the risk SNPs expressed lower levels of MLPH [35]. To investigate the mechanism by which *MLPH* contributes to prostate cancer risk, we performed functional studies in prostate organoids, 3D *in vitro* models of prostate epithelia homeostasis and function [36]. Cells within these *in vitro* cultures, derived from adult mouse prostate, differentiate and organize into the major cell types seen in the adult gland; inner luminal cells surrounded by basal cells.

We generated genetically modified prostate organoid clones using CRISPR/Cas9 and a sgRNA targeting a region of exon 3 of the mouse *Mlph* gene that codes for the RAB27a-binding domain. Sequencing genomic DNA from the organoid clones identified frameshift mutations in both *Mlph* alleles at the sgRNA target site

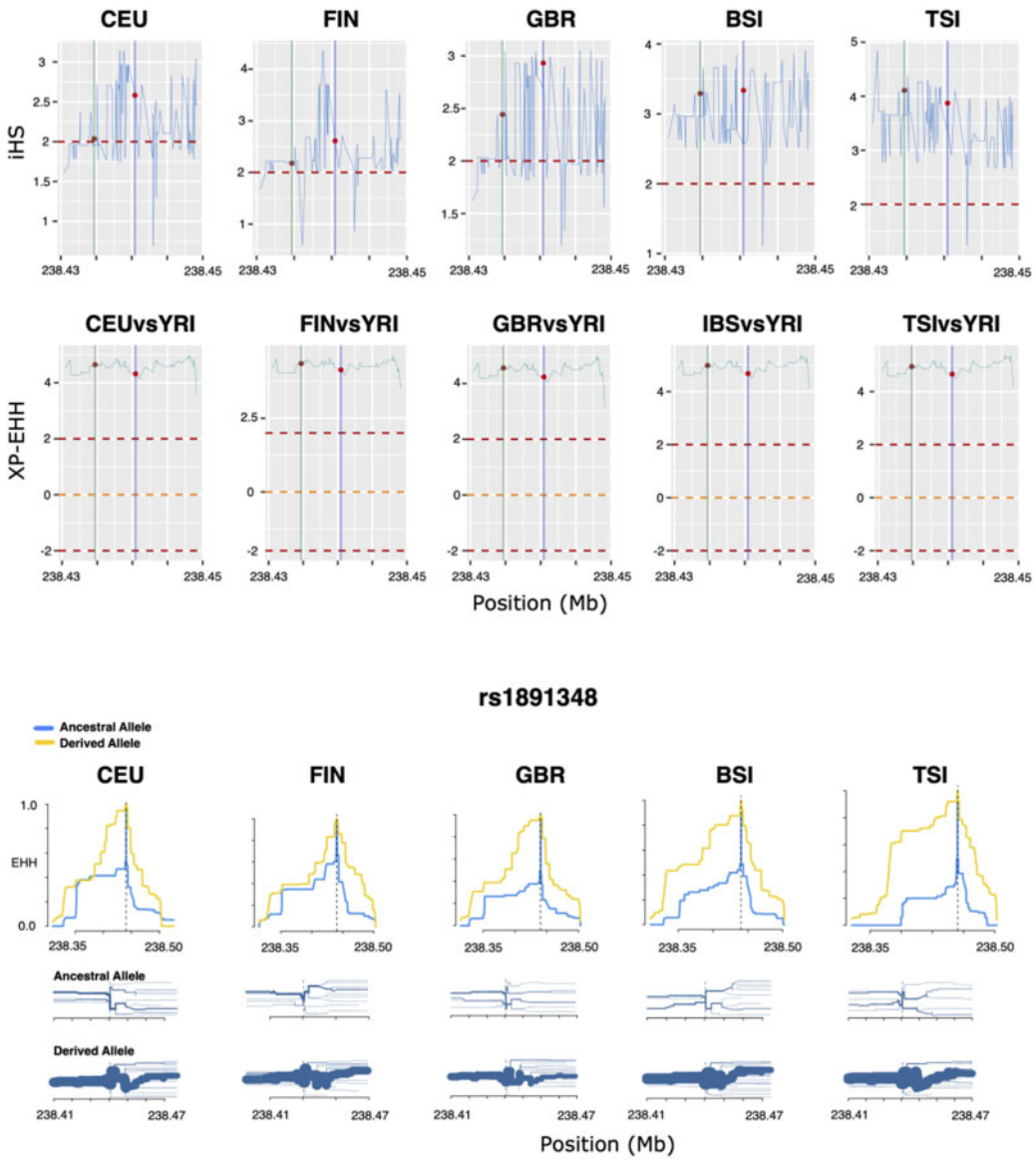

Figure 1. Selection on rs11891348 and rs11891426 identified by iHS and XP-EHH. The iHS scores are expressed as absolute value. Lines indicate iHS or XP-EHH scores for each SNP tested. The red dashed line marks the significant threshold. The genomic position of the SNP rs11891348 is indicated by the blue vertical line. The green vertical line indicates the genomic position of rs11891426. Red dot: rs11891348; brown dot: rs11891426.

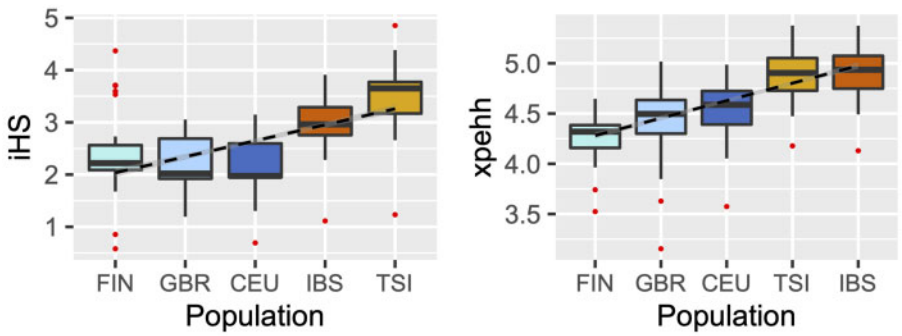

Figure 2. Selection strength acting on MLPH at different latitudes. Distribution of absolute iHS (left) and XP-EHH (right) scores for each European population. The dashed line represents a linear regression for trend with 95% of confidence interval (grey shadow).

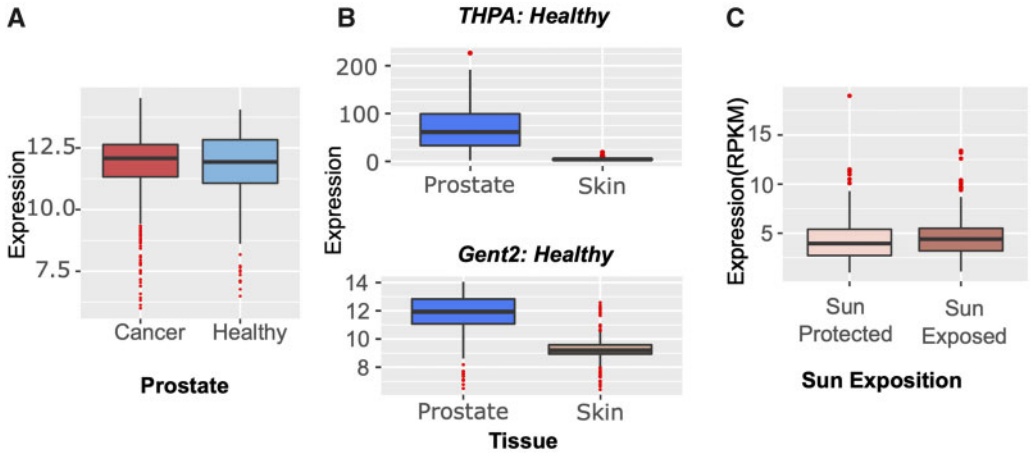

Figure 3. RNA expression of *MLPH*. (A) RNA expression in cancer and healthy prostate from Gent2 dataset (142 healthy and 620 cancer). No significant differences were found but the analysis might need more power. Sample size estimation was performed indicating an appropriate sample size of 3192 measurements. (B) RNA expression in healthy prostate and healthy skin. Top: GTEx (THPA) dataset; 106 and 607 measures in prostate and skin. Bottom: Gent2 dataset; 142 prostate and 319 skin. Expression measures are different between the two databases. (C) RNA expression in sun protected (250 measures) and exposed (357 measures) skin from GTEx (THPA) dataset.

(Supplementary Fig. S8C). *Mlph*-specific RNA transcripts were found in the mutant organoids suggesting that they did not undergo nonsense-mediated decay (Supplementary Fig. S8E–G). We also identified MLPH protein which may reflect the common event found in genes with CRISPR-induced frameshifts due to alternative splicing (Supplementary Fig. S8D) [37]. To confirm the generation of homozygous mutant clones, we analysed the *Mlph* mutant transcripts and found only mutant sequences at the target site (10/10 sequenced transcripts). Co-immunoprecipitation assay using MLPH and RAB27A antibodies was carried out to investigate if the protein seen in the mutant clones was functional. Western blot analysis on lysate from control and mutant organoid samples immunoprecipitated with MLPH antibody showed that RAB27A was bound to control but not to mutant MLPH protein (Fig. 4C). Actin, which is part of the melanosome transport complex, was also found to be preferentially associated with control and not mutant MLPH. These assays therefore show that we have generated prostate organoids that express mutant versions of MLPH protein that are functionally deficient.

Microscopic analysis of the mutant *Mlph* 3D cultures showed the presence of a higher number of organoids with lumens (Fig. 4A). *Mlph* mutants, tested by IHC assays, showed an increase in inner luminal cells (CK8 staining) and a decrease in outer basal cells (p63 staining) (Fig. 4B). This result was confirmed by qRT-PCR where *Mlph* mutants showed a significant increased expression of *CK18*, a luminal marker (*P* values <0.05) and a decreased expression of basal markers (*Trp63* and *CK5*; *P* values <0.05) (Fig. 4D). Mutant *Mlph* organoids showed the consistent presence of structures with no basal cells, which were not observed in control organoids. Luminal-only organoids are a property of prostate cancer cells compared to normal prostate and have been proposed to be a source of committed luminal progenitors that promote adenocarcinoma when grafted into mice [38]. Quantification of CK8 and p63 IHC staining confirmed that *Mlph* mutant samples contain more organoids with lumen, and include those that do not contain basal cells (Fig. 4E–G). Although we observed a reduction in the number of *Mlph* mutant organoid formation (about 49% decrease, Fig. 4H), cell viability assays on the organoid cultures revealed a slightly higher level of growth in the mutant (Fig. 4I). Further analysis showed mutant organoids were bigger than controls, in particular the ones containing lumens (Fig. 4J). The reduction in organoid formation may be due to the lower number of basal cells found in the mutant as these cells have been shown to be more efficient at forming organoids compared to luminal cells [36]. Therefore, our data are consistent with the loss of *Mlph* leading to an increase in luminal cells giving rise to larger lumen containing organoids.

MLPH functional loss promotes luminal cell resistance to androgen deprivation therapy

Prostate cells are dependent on the male hormone, androgens, and in the clinic, androgen withdrawal is the first line of therapy for prostate cancer patients. Inhibition of this pathway in patients is obtained using second-generation AR inhibitors such as enzalutamide. To investigate the response of *Mlph* mutant prostate cells to androgen deprivation, we treated organoid cultures with DHT, no DHT and no DHT plus enzalutamide (Fig. 5A and B). *Mlph* mutants and controls had similar sensitivities to androgen deprivation in cell viability assays with enzalutamide having a markedly higher effect on organoid growth than lack of DHT for both (Fig. 5D). Differences were observed in the type of organoids growing in these conditions. In control

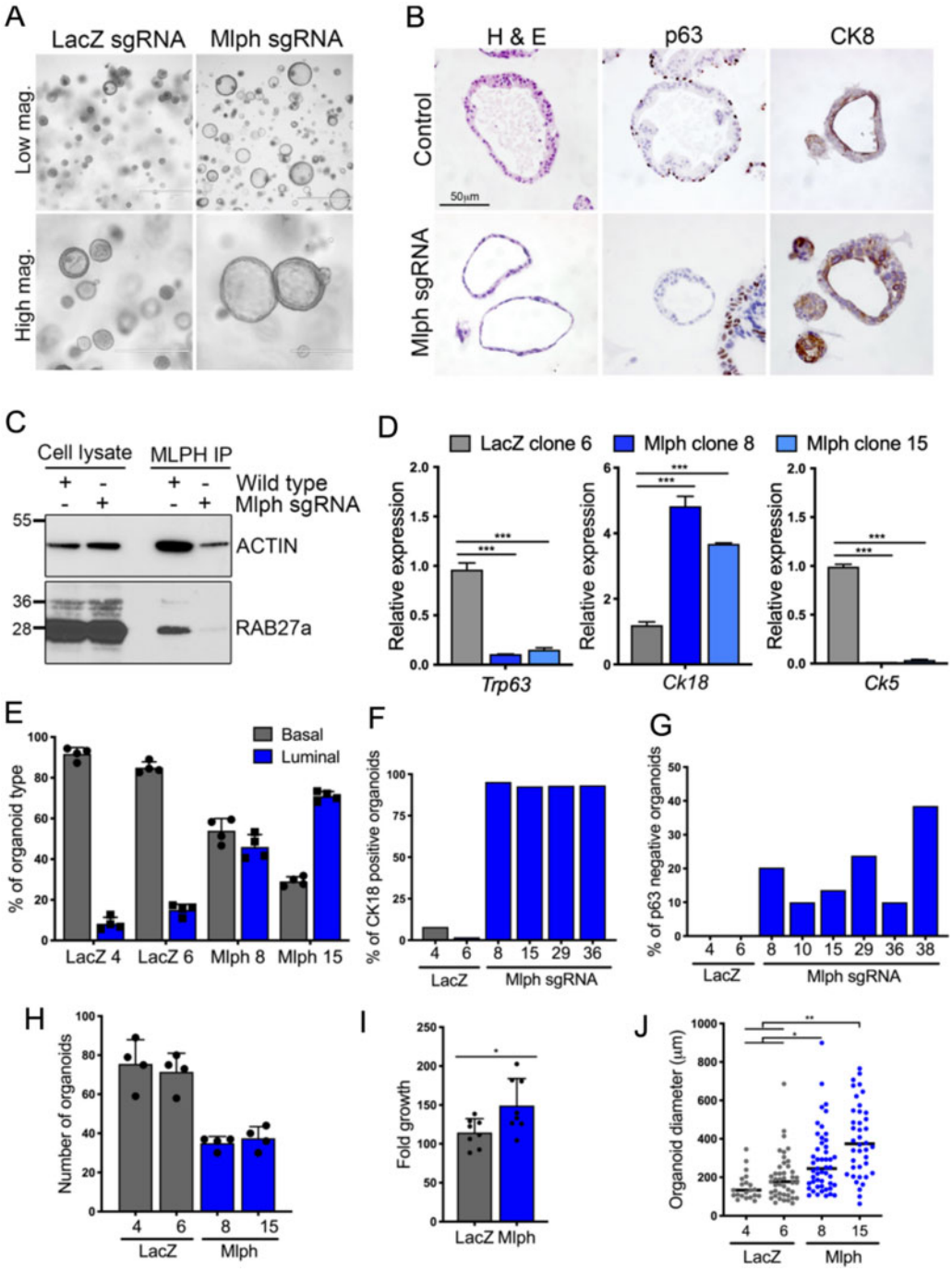

Figure 4. *Mlph* mutant prostate organoids contain more luminal cells. (A) Brightfield images of LacZ control and *Mlph* mutant organoids. Low magnification is 2× and high magnification is 4×. (B) Haematoxylin and Eosin (H & E) stain, p63 and CK8 immunohistochemistry on sections of control and *Mlph* mutant organoids. (C) Co-immunoprecipitation of MLPH with RAB27a and ACTIN in control and *Mlph* mutant organoids. Organoid lysates were immunoprecipitated with anti-MLPH antibody followed by Western blot analysis of RAB27a and ACTIN. (D) qRT-PCR of *Trp63*, *Ck5* and *Ck18* in control and mutant organoids. Mean and SD (error bars) are indicated. (E) the percentage of basal and luminal organoids formed from control and *Mlph* mutant cells. (F) the percentage of control and *Mlph* mutant organoids with CK18-positive cells, based on antibody stains. (G) the percentage of control and *Mlph* mutant organoids with no p63 positive cells, based on antibody stains. (H) Quantitation of the number of organoids formed from control and *Mlph* mutant cells. The number of organoids were counted from 4× images from four wells of each organoid line. (I) *Mlph* mutant and control organoid fold growth after seven days culture based on CellTitre Glo cell viability assay. (J) *Mlph* mutant and control organoid diameter. The significance of the data was analysed using a Student's t-test, and differences between two means with a P value <0.05 were considered significant. Error bars in the graphics represent the standard error of the mean.

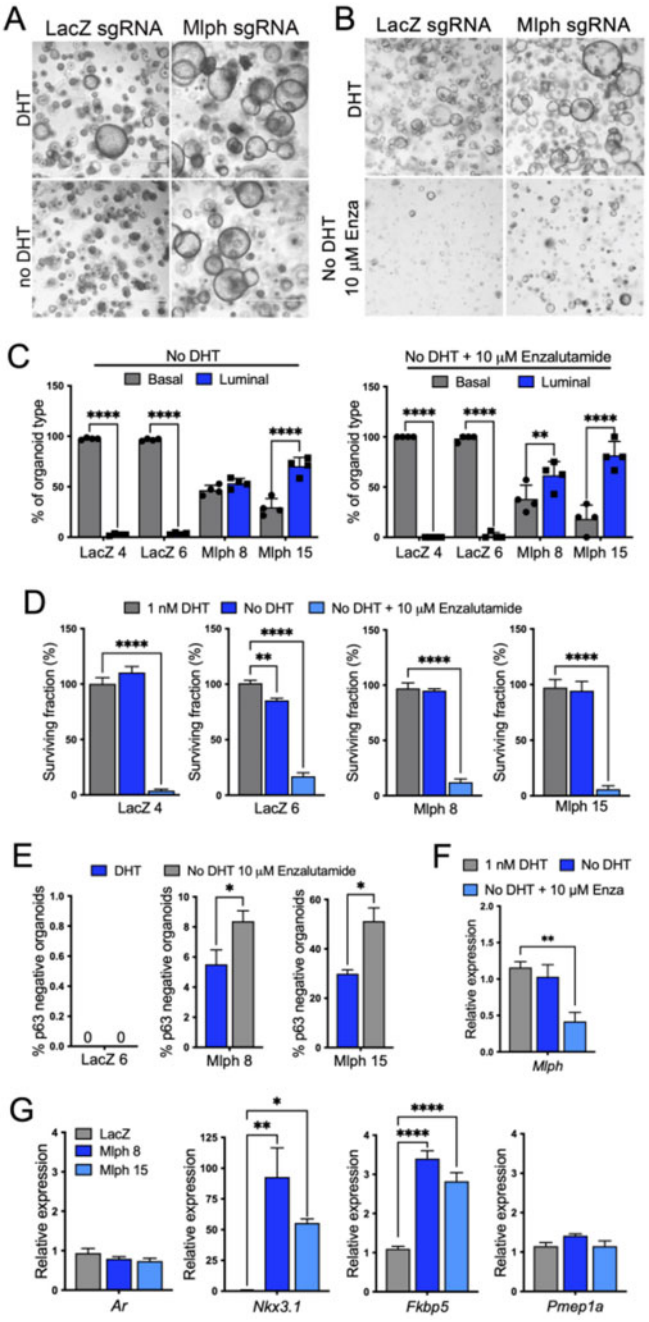

Figure 5. *Mlph* mutant organoids have an increased number of AR-independent luminal cells. Brightfield images of control and *Mlph* mutant organoids grown with DHT or (A) no DHT, and (B) no DHT and 10 μM enzalutamide. (C) The percentage of basal and luminal organoids formed from control and *Mlph* mutant prostate cells grown in no DHT or no DHT and 10 μM enzalutamide. (D) survival fraction of control and *Mlph* mutant cells grown with DHT, no DHT or no DHT and 10 μM enzalutamide as assayed by CellTitre Glo viability assay. Each sample normalized to DHT growth. (E) The number of control and *Mlph* mutant organoids grown in DHT or in no DHT and 10 μM enzalutamide with no p63 stain. (F) qRT-PCR of *Mlph* in control organoids grown in DHT, no DHT, or no DHT and 10 μM enzalutamide. Mean and SD (error bars) are indicated. (G) qRT-PCR of *Ar, Nkx3.1, Fkbp5 and Pmep1a* in control and *Mlph* mutant organoids. Mean and SD (error bars) are indicated.

cultures, both enzalutamide treatment and lack of DHT gave rise to mostly organoids without lumens, with solid basal type organoid growth being promoted (compare LacZ4 and LacZ6 cultures in Fig. 5C to 4E). In contrast, in the *Mlph* mutants luminal type organoids were present in both androgen deprivation treatments. The difference was particularly prominent in the enzalutamide-treated cultures as lumen containing organoids were totally absent in control samples (Fig. 5C). IHC staining for the basal cell marker p63 within mutant samples (DHT treated and no DHT plus enzalutamide) showed enzalutamide-treated samples being enriched with luminal-only organoids (Fig. 5E). These data show that loss of *Mlph* function leads to the increased survival of prostate luminal cells in androgen deprivation conditions. To further investigate the relationship between MLPH and AR signalling, we analysed the expression of MLPH in prostate organoids grown without DHT and in the presence of enzalutamide. MLPH expression was found to be reduced in the enzalutamide treated sample compared to controls grown in DHT (Fig. 5F). Androgen target gene expression analysis on mutant MLPH organoids showed an increase in mutant samples compared to control samples in some target genes, namely *Nkx3.1* and *Fkbp5*, but not others, such as *Ar* and *Pmep1a* (Fig. 5G).

DISCUSSION

In the present study, we dissected the genetic architecture of a limited set of loci associated with breast/prostate cancer susceptibility to determine if any had signatures of positive selection. Our analysis showed no evidence of selection in almost all cases. We conclude that the high frequency of these allelic variants in European populations most likely reflects neutral drift or founder effects. In contrast, we found strong evidence for recent positive selection in *MLPH* for two common SNPs rs11891426 [35] and rs11891348 [15] as well as for the whole *MLPH* gene. A recent study has also found signatures of positive selection in *MLPH* in the CEU population [39]. When we explored which of the two alleles underwent selective sweeps, we found that in both SNPs, the derived and protective T alleles were the ones with long-range linkage disequilibrium at the core SNP. The T alleles of both SNPs are therefore significantly correlated. Although a long-range haplotype carrying the derived TT alleles is present in high frequency in all European populations examined here, it is worth noting the presence of different really low frequency long range haplotypes carrying both ancestral and derived alleles at the core haplotype (Fig. 1 and Supplementary Fig. S5B). This haplotype diversity might suggest the occurrence of few recombinations and thus probable recent and/or ongoing selection.

The historical, adaptive logic of a positively selected allele can be difficult to determine. The *MLPH* alleles we found to

carry the long-range haplotype, reduce the risk (OR : 0.9) of prostate cancer but, as lethal prostate cancer is almost entirely post-reproductive, it is very unlikely that cancer itself provided the selective pressure or adaptive advantage [5]. This most likely derived from an impact of *MLPH* alleles on skin pigmentation responses.

The *MLPH* gene encodes for the MLPH protein involved in melanosome transport [33]. The MLPH protein forms a ternary complex with the small Ras-related GTPase Rab27a and the motor protein myosinVa (Rab27a-Mlph-MyoVa) where MLPH acts as a tether to enhance the transfer of melanosomes from melanocytes to the adjacent keratinocyte providing pigment needed for hair, skin and eye colouring [40]. In humans, mutations in each of the above three genes lead to different types of Griscelli syndrome, a hypopigmentation disorder [41]. When skin is exposed to the sun, the expression of *MLPH* is significantly higher than the sun-protected skin suggesting *MLPH* expression associated with UV irradiation response. Transfer of melanosomes from melanocytes to keratinocytes enables darker pigmentation and plays an essential role in protecting the skin from UV irradiation [42]. Higher expressed *MLPH* is therefore associated with darker phenotype in sheep [43] and rabbits [44]. In Europeans, pigmentation phenotypes were found to reflect the geographic allocation, lighter in the north and darker in the south [45]. We found a north-south latitudinal cline for selection strength acting on *MLPH*, possibly reflecting evolutionary adaptation to latitude-dependent UV levels and a role in skin darkening/tanning might be plausible. The stronger selection pressure acting in the south may reflect an evolutionary drive for higher UV protection, which incidentally impacted prostate cancer risk. Sun exposure was found to inversely correlate with prostate cancer incidence [46] and a past research showed a north-south trend for prostate cancer mortality, with lower rates in the south with higher UV levels [47]. Environmental conditions related to different UV levels might have exerted on European ancestors a selective pressure and likely those TT (rs11891426 and rs11891348) carriers might have had a stronger response to the sun exposure.

Despite its function in skin, *MLPH* is not exclusively expressed in the skin, but is highly expressed in healthy and malignant prostate cancer. Within prostatic tumours, lower levels of *MLPH* are associated with more aggressive disease [48]. Similarly, in breast cancer, the more malignant oestrogen negative tumours express lower levels of *MLPH* than oestrogen positive tumours implying hormonal regulation. Rs11891426 protective allele carriers express higher *MLPH* levels in prostate tumours [35].

Our functional studies on *MLPH* loss in cultured organoids of mouse prostate tissue suggest a possible rationale of the impact of *MLPH* alleles on prostate cancer risk and aggressiveness. Loss of *MLPH* function in prostate organoids promotes luminal differentiation and growth even in androgen deprivation conditions. In addition, our studies show the presence of luminal-only structures specifically in mutant cultures, which are more resistant to enzalutamide treatment. These structures have been found to contain luminal precursors and are preferentially found in cultures from aggressive prostate tumour tissue with loss of tumour suppressors *Tp53* and *Pten* [38]. Studies in mice have shown that luminal cells are more sensitive to neoplastic transformation by the loss of *Pten* than basal cells of the prostate [49].

The risk allele for *MLPH* expresses lower levels of protein and one possibility is that this, consistent with our *Mlph* mutant studies, increases the number of luminal progenitor/stem cells—the likely target cell population for drivers of clinical prostate cancer [50], which is mostly luminal in phenotype. The selected SNP alleles reside within the androgen-binding site of *MLPH*. Consistent with this, we did observe a decrease in MLPH expression in prostate organoids following androgen deprivation therapy. Lack of MLPH led to changes in the expression of some androgen target genes. Therefore, our data suggest an interplay between androgen signalling and *MLPH* expression and a luminal phenotype, which may change with neoplastic transformation and patient treatment.

SUPPLEMENTARY DATA

Supplementary data is available at EMPH online.

ACKNOWLEDGEMENTS

The authors thank Mitsunori Fukuda for his anti Slac2-a antibody and Professor Ros Eeles for data summaries on SNP acquisition in prostate cancer and breast cancer, respectively.

AUTHOR'S CONTRIBUTIONS

L.E. analysed the data, carried out the statistical and bioinformatics analysis, interpreted the results and contributed to devise the study. J.C.F. performed the functional study with contribution from G.S.R and J.N. A.S designed, supervised and interpreted the functional experiments. M.G. conceived and planned this study, interpreted the results and supervised all aspects of this work. L.E. and M.G. wrote the manuscript with contributions from A.S. and input from all authors.

FUNDING

This work was supported by a Wellcome Trust award to the Centre for Evolution and Cancer (105104/Z/14/Z) and The Institute of Cancer Research, London (M.G.). J.C.F. and G.S.R. were funded by a Prostate Cancer UK grant (RIA17-ST2-01).

Conflict of interest: None declared.

REFERENCES

1. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer* 2017; **17**:692–704.
2. Van de Haar J, Canisius S, Yu MK *et al.* Identifying epistasis in cancer genomes: a delicate affair. *Cell* 2019; **177**:1375–83.
3. Rafnar T, Sulem P, Stacey SN *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* 2009; **41**:221–7.
4. Bien SA, Peters U. Moving from one to many: insights from the growing list of pleiotropic cancer risk genes. *Br J Cancer* 2019;**120**:1087–9.
5. Medawar PB. An Unsolved Problem of Biology. London: H.K. Lewis,1952.
6. Williams GC. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 1957; **11**:398–411.
7. Greaves M. *Cancer: The Evolutionary Legacy*. New York: Oxford University Press, 2001.
8. Erdei E, Torres SM. A new understanding in the epidemiology of melanoma. *Expert Rev Anticancer Ther* 2010;**10**:1811–23.
9. Zeron-Medina J, Wang X, Repapi E *et al.* A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell* 2013;**155**:410–22.
10. Kumar S, Liu L. No positive selection for G allele in a p53 response element in Europeans. *Cell* 2014;**157**:1497–9.
11. Ding Y, Larson G, Rivas G *et al.* Strong signature of natural selection within an FHIT intron implicated in prostate cancer risk. *PLoS One* 2008;**3**:e3533.
12. Gaston KE, Kim D, Singh S *et al.* Racial differences in androgen receptor protein expression in men with clinically localized prostate cancer. *J Urol* 2003;**170**:990–3.
13. Smith KR, Hanson HA, Mineau GP *et al.* Effects of BRCA1 and BRCA2 mutations on female fertility. *Proc Biol Sci* 2012;**279**:1389–95.
14. Ferla R, Calò V, Cascio S *et al.* Founder mutations in BRCA1 and BRCA2 genes. *Ann Oncol* 2007;**18 Suppl 6**:vi93–8.
15. Amin Al Olama A, Dadaev T, Hazelett DJ, *et al.*; UK ProtecT Study Collaborators. Multiple novel prostate cancer susceptibility signals identified by fine-mapping of known risk loci among Europeans. *Hum Mol Genet* 2015;**24**:5589–602.
16. Sabeti PC, Schaffner SF, Fry B *et al.* Positive natural selection in the human lineage. *Science* 2006;**312**:1614–20.
17. Auton A, Brooks LD, Durbin RM, *et al.*; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
18. Voight BF, Kudaravalli S, Wen X *et al.* A map of recent positive selection in the human genome. *PLoS Biol* 2006;**4**:e72.
19. Sabeti PC, Varilly P, Fry B, *et al.*; International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;**449**:913–8.
20. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 2014;**31**:2824–7.
21. Sabeti PC, Reich DE, Higgins JM *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002;**419**: 832–7.
22. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**:585–95.
23. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics* 2000;**155**:1405–13.
24. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE *et al.* PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* 2014;**31**:1929–36.
25. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019.
26. Uhlén M, Fagerberg L, Hallström BM *et al.* Proteomics. Tissue-based map of the human proteome. *Science* 2015;**347**:1260419.
27. Park S-J, Yoon B-H, Kim S-K *et al.* GENT2: an updated gene expression database for normal and tumor tissues. *BMC Med Genomics* 2019;**12**:101.
28. Drost J, Karthaus WR, Gao D *et al.* Organoid culture systems for prostate epithelial and cancer tissue. *Nat Protoc* 2016;**11**:347–58.
29. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 2014;**11**:783–4.
30. Enattah NS, Sahi T, Savilahti E *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002;**30**:233–7.
31. Bersaglieri T, Sabeti PC, Patterson N *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004;**74**:1111–20.
32. Cadzow M, Boocock J, Nguyen HT *et al.* A bioinformatics workflow for detecting signatures of selection in genomic data. *Front Genet* 2014;**5**:293.
33. Oberhofer A, Spieler P, Rosenfeld Y *et al.* Myosin Va's adaptor protein melanophilin enforces track selection on the microtubule and actin networks in vitro. *Proc Natl Acad Sci U S A* 2017;**114**:E4714–23.
34. Dadaev T, Saunders EJ, Newcombe PJ, *et al.*; The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat Commun* 2018;**9**:2256.
35. Bu H, Narisu N, Schlick B *et al.* Putative prostate cancer risk SNP in an androgen receptor-binding site of the melanophilin gene illustrates enrichment of risk SNPs in androgen receptor target sites. *Hum Mutat* 2016;**37**:52–64.
36. Karthaus WR, Iaquinta PJ, Drost J *et al.* Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* 2014; **159**:163–75.
37. Smits AH, Ziebell F, Joberty G *et al.* Biological plasticity rescues target activity in CRISPR knock outs. *Nat Methods* 2019;**16**:1087–93.
38. Agarwal S, Hynes PG, Tillman HS *et al.* Identification of different classes of luminal progenitor cells within prostate tumors. *Cell Rep* 2015;**13**:2147–58.
39. Lachance J, Berens AJ, Hansen MEB *et al.* Genetic hitchhiking and population bottlenecks contribute to prostate cancer disparities in men of African descent. *Cancer Res* 2018;**78**:2432–43.
40. Sckolnick M, Krementsova EB, Warshaw DM *et al.* More than just a cargo adapter, melanophilin prolongs and slows processive runs of myosin Va. *J Biol Chem* 2013;**288**:29313–22.
41. Ménasché G, Ho CH, Sanal O *et al.* Griscelli syndrome restricted to hypopigmentation results from a melanophilin defect (GS3) or a MYO5A F-exon deletion (GS1). *J Clin Invest* 2003;**112**:450–6.
42. Natarajan VT, Ganju P, Ramkumar A *et al.* Multifaceted pathways protect human skin from UV radiation. *Nat Chem Biol* 2014;**10**: 542–51.
43. Fan R, Xie J, Bai J *et al.* Skin transcriptome profiles associated with coat color in sheep. *BMC Genomics* 2013;**14**:389.
44. Demars J, Iannuccelli N, Utzeri VJ *et al.* New insights into the melanophilin (*MLPH*) gene affecting coat color dilution in rabbits. *Genes* 2018;**9**:430.

45. Candille SI, Absher DM, Beleza S *et al.* Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. *PLoS One* 2012;**7**:e48294.
46. Taksler GB, Cutler DM, Giovannucci E *et al.* Ultraviolet index and racial differences in prostate cancer incidence and mortality. *Cancer* 2013; **119**:3195–203.
47. Hanchette CL, Schwartz GG. Geographic patterns of prostate cancer mortality. Evidence for a protective effect of ultraviolet radiation. *Cancer* 1992;**70**:2861–9.
48. Mancuso N, Gayther S, Gusev A, *et al.*; PRACTICAL Consortium. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun* 2018;**9**:4079.
49. Choi N, Zhang B, Zhang L *et al.* Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer Cell* 2012;**21**:253–65.
50. Zhang D, Zhao S, Li X *et al.* Prostate luminal progenitor cells in development and cancer. *Trends Cancer* 2018;**4**:769–83.