Directional and balancing selection on proprotein convertase subtilisin/kexin type 6 (PCSK6)

Sonja Kuderer Corresp., [1] , **Martin Fieder** [1] , **Sylvia Kirchengast** [1]

[1] Deparment of Anthropology, Universität Vienna, Vienna, Austria

Corresponding Author: Sonja Kuderer
Email address: sonja.kuderer@gmail.com

The selective pressures challenging the human population leave genetic signatures. This study was designed to detect positive and balancing selection on the proprotein convertase subtilisin/kexin type 6 (PCSK6) gene by four different approaches (FDIST, BayeScan, EHH, iHS). The outlier programs FDIST and BayeScan found 27 overlapping single nucleotide polymorphisms (SNPs) under selection, of which twelve are under strong positive directional selection and 15 under balancing selection. In the EHH analysis, all SNPs detected under positive selection show a slow decay of the derived extended haplotype. The integrated haplotype scores (iHS) for eight loci under positive selection were found to be larger than one. Based on these results, we provide a short overview of PCSK6-related mechanisms potentially associated with the detected positive or balancing selection patterns. Positive selection possibly reflects the key role of PCSK6 in tumorigenesis, in nociception, in rheumatoid arthritis and osteoarthritis, in the cardiovascular circulatory as well as in blood pressure regulation. Balancing selection acting on the PCSK6 locus might be explained by its involvement in vertebral left-right asymmetry and therefore in human handedness. Whereas until now the exact nature of the selective pressures acting on PCSK6 remained incompletely understood, the results of this study provide for the first time insights into the enzyme's evolutionary significance owing to its functional manifoldness.

Directional and balancing selection on proprotein convertase subtilisin/kexin type 6 (PCSK6)

Sonja Kuderer[1], Sylvia Kirchengast[1], Martin Fieder[1]

[1]Department of Anthropology, University of Vienna, Vienna, Austria

Corresponding author

Sonja Kuderer[1]

Althanstrasse 14, 1090 Vienna, AUSTRIA

Email address: sonja.kuderer@gmail.com

Abstract

The selective pressures challenging the human population leave genetic signatures. This study was designed to detect positive and balancing selection on the proprotein convertase subtilisin/kexin type 6 (PCSK6) gene by four different approaches (FDIST, BayeScan, EHH, iHS). The outlier programs FDIST and BayeScan found 27 overlapping single nucleotide polymorphisms (SNPs) under selection, of which twelve are under strong positive directional selection and 15 under balancing selection. In the EHH analysis, all SNPs detected under positive selection show a slow decay of the derived extended haplotype. The integrated haplotype scores (iHS) for eight loci under positive selection were found to be larger than one. Based on these results, we provide a short overview of PCSK6-related mechanisms potentially associated with the detected positive or balancing selection patterns. Positive selection possibly reflects the key role of PCSK6 in tumorigenesis, in nociception, in rheumatoid arthritis and osteoarthritis, in the cardiovascular circulatory as well as in blood pressure regulation. Balancing selection acting on the PCSK6 locus might be explained by its involvement in vertebral left-right asymmetry and therefore in human handedness. Whereas until now the exact nature of the selective pressures acting on PCSK6 remained incompletely understood, the results of this study provide for the first time insights into the enzyme's evolutionary significance owing to its functional manifoldness.

Introduction

Human-specific phenotypic and behavioural variability can be at least partly explained by natural selection. Selective pressures varying concerning time, strength and mode of selection have challenged human populations (Karlsson et al., 2014). An investigation of these natural selection processes acting on particular genetic loci is therefore suited to provide insights into the functional relevance of genes.

Natural selection describes the tendency of beneficial traits to increase their frequency in populations over time and affects specific functionally important sites in the genomes (Darwin & Wallace, 1858). Positive selection leads to an increase in the frequency of genetic variants, which provide improved survival skills and/or a higher fertility rate. A frequent substrate of positive selection, as revealed by numerous studies in different taxa, are genes for reproduction, dietary adaption, physical appearance, nociception, brain development or behavioural traits (Vallender & Lahn, 2004). Whilst negative selection eliminates existing disadvantageous variants within populations, balancing or diversified selection favours genetic diversity by maintaining multiple alleles at particular loci (Karlsson et al., 2014). Common mechanisms of balancing selection are negative frequency-dependent selection (e.g. host-pathogen interaction, Leffler et al., 2013), spatial or temporal habitat heterogeneity, sexual antagonism (Delph & Kelley, 2014) or heterozygote advantage.

Human handedness, the most prominent laterality phenomenon along the human body, shows a genetic heritability and a worldwide variation across populations (Raymond & Pontier, 2004). From an evolutionary point of view, the existence of a

handedness polymorphism cannot simply be explained by a neutral handedness trait (Faurie & Raymond, 2013). Hence, due to the genetic basis of handedness, proposed candidate genes might be regarded being subjected to natural selection.

One candidate gene for handedness is the proprotein convertase subtilisin/kexin type 6 (PCSK6), which is located on the long arm of the human chromosome 15 (15q26; Kiefer et al., 1991). Brandler et al. (2013) proved a relation of PCSK6 *rs7182874* with relative hand skill at genome-wide significance among people with reading disability. In addition, Scerri et al. (2011) found a connection between two intronic single point mutations in PCSK6 (*rs9806256, rs11855415*) and relative hand skill in individuals with reading disability in a genome-wide association study (genome-wide significance level with a p-value of $1.99e^{-8}$). Amongst these individuals, carriers of the minor PCSK6 *rs11855415* allele revealed a significantly greater relative right-hand skill compared to the more frequent wild-type allele. However, *rs11855415* and *rs7182874* were not associated with hand skill in a general population unaffected by reading disability (Brandler et al., 2013; Scerri et al., 2011). Furthermore, a significant link between the hand preference and an intron 33bp variable number of tandem repeat (VNTR) polymorphism in PCSK6 (*rs10523972*) was discovered (Arning et al., 2013). Therefore it is assumed, that PCSK6 rather than determining the direction of handedness is involved in the manifestation of a certain degree of handedness (Arning et al., 2013).

PCSK6 belongs to the serine endoprotease superfamily, which catalyse the cleavage of a wide range of substrates including secreted growth factors and extracellular pathogens (Seidah & Prat, 2012). The PCSK6 locus consists of at least

250 kb (bp 101,840,818 to bp 102,065,405), contains 25 exons and is therefore the largest mammalian kexin-like proprotein convertase known to date (Tsuji et al., 1997). The enzyme has 20 splice variants (Tsuji et al., 1997) and seven isoforms which vary in size and the 3' coding sequence (Tsuji et al., 1994).

The PCSK6 protein exhibits a general organization starting from the N-terminus with a signal peptide, a propeptide, a subtilisin-like catalytic domain with active site residues and a β-barrel P or homo B domain; the latter is thereby essential for the enzyme's proteolytic activity (Li et al., 2004). In addition, PCSK6 has a long cysteine-rich region at the carboxy-terminus, which varies among the particular isoforms and provides an explanatory approach for their differing intracellular localizations and functions (Tsuji et al., 1997). This C-terminal cysteine-rich domain confers properties to PCSK6 to attach to tissue inhibitors of metalloproteinases as well as to the cell surface (Nour et al., 2005). Thus, PCSK6 is a secreted, heparin-binding and extracellular matrix-anchored protein that is localized at the cell surface (Seidah & Prat, 2012).

The PCSK6 convertase is initially synthesized as an inactive zymogen. After the primary autocatalytic processing event in the endoplasmic reticulum, a heterodimer is formed from the inhibitory prosegment and the rest of the molecule. This heterodimer is subsequently transported to the trans-Golgi network, where the catalytic activity of the enzyme is acquired after a second autocatalytic cleavage event (Nour et al., 2004; Seidah & Prat, 2007). Hence, PCSK6 is synthesized as a proenzyme which undergoes a proteolytic removal of the N-terminal propetide (Nagahama et al., 1998).

PCSK6 exhibits a unique cellular distribution in various tissues such as the liver, lung, uterine endometrium, gut, neuroendocrine tissue or the epidermis. An especially high level of PCSK6 is present in the corpus callosum (Diaz et al., 2008) and the nervous as well as the pituitary system (Nagahama et al., 1998). At its target sites this calcium-dependent serine endoprotease processes precursor proteins into biologically active products. As PCSK6 cleaves substrates at a single or paired basic residue within the (K/R)-$(X)_n$-(K/R) consensus site, with n being 0, 2, 4 or 6 and X any amino acid (Turpeinen et al., 2013) it is also known as paired basic amino acid cleaving enzyme 4 (PACE 4). However, to avoid confusion in the naming systems, we adhere to the PCSK6 naming convention.

Due to its localization to the cell membrane, PCSK6 is mainly responsible for processing membrane-associated substrates (Nour et al., 2005) and might be regulated by extracellular signals (Bassi et al., 2010). Transforming growth factor beta (TGFbeta)-related proteins (Constam & Robertson, 2000), proalbumin (Mori et al., 1999) and luteinizing hormone (Wang et al., 2014) are found among the well-documented interaction partners of PCSK6. Further suggested PCSK6 targets include the growth factors Nodal and Lefty as well as metalloproteinases such as ADAMTS4 (Tortorella et al., 2005).

PCSK6 is moreover responsible for the proteolytic maturation of many growth factors, hormones, neuropeptides and zymogens (Seidah et al., 2008). In general, the PCSK6 enzyme is involved in manifold biological processes which vary from tumour progression (D'Anjou et al., 2011), pathogenesis of osteoarthritis (Tortorella

et al., 2005) to the control of diastolic blood pressure (Li et al., 2004). PCSK6 also occupies a considerable role in biological mechanisms regulating the establishment of normal brain lateralization and handedness. It apparently has this control function during the anterior central nervous system patterning and the formation of left-right asymmetry in vertebrates (Constam & Robertson, 2000). More specifically, PCSK6 acts upstream of Nodal, Lefty1 and Lefty2 and thereby influences the balance between Nodal and bone morphogenic protein (BMP) signalling in the lateral plate, which is critical for left-right axis formation (Constam & Robertson, 2000). Therefore, mice embryos with a lacking or a disrupted PCSK6 locus develop situs ambiguous, a rare congenital defect in which the major visceral organs are distributed abnormally within the chest and abdomen, and also show complex craniofacial malformations as well as asymmetry defects such as heterotaxia (Constam & Robertson, 2000). 25% of PCSK6 knock-out mice die at embryonic day 14 with severe cardiac malformation and bone morphogenic defects (Constam & Robertson, 2000).

By affecting the fitness of the organism, functionally important genomic loci usually show a high evolutionary selection pressure (Nielsen, 2005). This study therefore applies four different approaches to identify if there exist natural selection signatures acting on the PCSK6 locus. It is aimed to test if positive directional selection and/or balancing selection has been acting on PCSK6. To discover this locus-specific selection and adaptive divergence, two separate outlier analyses of DNA polymorphisms are conducted (Nielsen, 2005; Beaumont & Nichols, 1996; Foll & Gaggiotti, 2008). In addition, the “extended haplotypes” and the “integrated haplotypes” of these SNPs on the PCSK6 gene are calculated (Voight et al., 2006; Sabeti et al., 2002). Hence, these numerous applied statistical evaluations enable a

reduction of false-positive results and reveal whether the PCSK6 polymorphisms under evolutionary selection markedly differ from neutral expectations.

Materials & Methods

Genomic data. We tested the genetic material of 14 analysed human populations (n = 1092 individuals; Table 1) from the 1000 Genome Browser (http://browser.1000genomes.org/index.html/) to retrieve the coordinates of the PCSK6 gene. By computing the data with SPSmart engine v5.1.1 (http://spsmart.cesga.es/engines.php; Amigo et al., 2008) we obtained the genotypes of SNPs on PCSK6 with a minor allele frequency (MAF) of 5% (N=735 SNPs). We converted the raw genotype of these to “convert format” and further translated it to i) “genpop format” (LOSITAN) and ii) “geste format” (BayeScan) using PGD-Spider (http://www.cmpg.unibe.ch/software/PGDSpider/; Lischer & Excoffier, 2012). The probability of false positives was reduced by applying two different programs to detect selection. Both, LOSITAN/FDIST (i) and BayeScan (ii) refer to F_{ST} outlier approaches and detect SNPs under selection by a pattern deviating from neutrality. In general, loci under balancing selection show unusually low levels of genetic differentiation (i.e., $F_{ST} < 0.05$), whereas loci under diversifying selection reveal unusually high levels of differentiation (i.e., $F_{ST} > 0.25$) among populations (Balloux & Lugon-Moulin, 2002).

LOSITAN (FDIST) approach. We conducted an F_{ST} outlier analysis using LOSITAN Selection Workbench (http://popgen.net/soft/lositan/; Beaumont & Nichols, 1996). This approach is based on the expected distribution of Wright's inbreeding coefficient F_{ST} vs. H_E under an island model of migration with neutral markers. Outlier loci,

which show selection, can be assumed to have an excessively high or low F_{ST} value. We simulated the neutral distribution of F_{ST} with 100,000 iterations at a significance of P < 0.005 (P (Simulation F_{ST} < sample F_{ST}) to reduce the number of false positives (Antao et al., 2008). We performed the runs using the infinite allele model. To avoid overestimating the percentage of outlier loci, we computed a multiple testing correction based on false discovery rates (FDR) of LOSITAN.

BayeScan approach. The BayeScan 2.1 programme (http://cmpg.unibe.ch/software/bayescan/; Foll & Gaggiotti, 2008) enables by the application of Bayesian statistics a second F_{ST}-based method to detect outlier loci under selection. As these show a genetic differentiation among populations in varying environments, the estimation of the posterior probability of a locus being under selection can be calculated by a reversible-jump Markov Chain Monte Carlo (MCMC) approach. This process is based on two alternative models, of which one includes the effect of selection and the other excludes this factor. By calculating a logistic regression, the F_{ST}-coefficients of a selection process can be separated into a population-specific component (β_i) shared by all loci, and a locus-specific component (α_i) shared by all populations. A diversifying selection is associated with significant positive α_i values, whereas significant negative α_i values indicate a balancing or purifying selection. When a locus is presumably subject to selection (P ($\alpha_i \neq 0$)) the according posterior probability is estimated from the MCMC output by counting the number of times α_i is included in the model. After 20 pilot runs of 10,000 iterations each and an initial burn-in of 1,000,000 steps each, we ran the chain for 100,000 updating steps. The programme defines a q-value, which is the FDR analogue to the P value. We chose a log10 Bayes Factor (BF) ≥ 1.5 (i.e., Jeffreys' scale of evidence for BF = "very

strong") as the main criterion, which corresponds to a range of q-values from 0.03-0.01. These values should provide very strong evidence that a locus is under selection (http://cmpg.unibe.ch/software/bayescan/). To control the FDR and to generate the graphical outputs the R-script by the BayeScan programme is utilised.

Regression tree. To identify the SNPs separating the 14 analysed populations, we generated a regressing tree using the R-library r-part (R version 3.0.1). We assessed information on the specific SNPs via Ensembl (http://www.ensembl.org/).

Extended haplotype homozygosity. We verified the detection of positive selection by LOSITAN and BayeScan with extended haplotype homozygosity (EHH). EHH is the probability that two chromosomes carrying the core haplotype of interest are identical by descent for the entire interval of the region under investigation (Sabeti et al., 2002). This method therefore detects the transmission of an extended haplotype without recombination events. EHH detects only "recent" selective sweeps, respectively incomplete selective sweeps (Voight et al., 2006) within a population. The approach was implemented in "selscan" (https://github.com/szpiech/selscan/blob/master/manual/selscan-manual.pdf) based on the haplotypes estimated by SHAPEIT (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html). Due to the processing of very large amounts of data (according to the computational power available) we exemplarily calculated EHH only for the YRI, CEU, CHB and MXL populations. Although we are aware that due to population admixture in the MXL population results could be blurred, it was our prime interest to include exemplarily the data sets of four large populations located on different continents.

Integrated haplotype scores. We also included the iHS scores (integrated haplotype scores) for the YRI, CEU, CHB and MXL population provided by hapbin (http://datashare.is.ed.ac.uk/handle/10283/714) and the UCSC genome browser (https://genome.ucsc.edu/) for the SNPs under balancing and directional positive selection. iHS are a standardized measure that describe the size of extended haplotype homozygosity (EHH) at a genetic locus (SNP) along to the ancestral allele relative to the derived allele (Voight et al., 2006). Therefore analyses of iHS scores have the advantage that the signal strength of different SNPs is independent of the allele frequencies, which makes these values directly comparable (Voight et al., 2006).

Results

PCSK6 regression tree. In the regression tree, *rs9920839* is on the primary node (Fig. 1) and firstly separates the populations. *Rs9920839* was also found by both LOSITAN and BayeScan under directional positive selection (Fig. 2, Fig. 3 and Table 2). The populations are further divided by *rs6598455* and *rs8035554* (both at the second nodal point layer) as well as *rs2412067* and *rs56278789* (both at the third nodal point layer). Besides *rs28641892, rs874940, rs61360191, rs3784496, rs2005063* and *rs11852754* (Fig. 1) additionally split the populations into several minor subgroups. Also, both LOSITAN and BayeScan identified *rs3784496* under directional positive selection, which separates the ASW population from European populations such as FIN and TSI (Fig. 2, Fig. 3 and Table 2).

Outlier detection methods. LOSITAN identified 18 SNPs under directional positive selection ($p < 0.005$ – Fig. 2): *rs4965825, rs1871974, rs900414, rs3784519,*

rs3784518, rs3784515, rs28360762, rs28469755, rs12592414, rs7167668, rs9920839, rs6598455, rs4965834, rs3784496, rs3825902, rs1000914, rs8043067, rs11247300 and 22 SNPs under balancing selection: *rs11247297, rs11634270, rs11853956, rs11854179, rs11855154, rs11855415, rs12915828, rs1495271, rs2047216, rs28529205, rs34548635, rs36010356, rs4965869, rs55794955, rs62027207, rs6598477, rs71414401, rs71416212, rs7166089, rs7168817*. BayeScan detected 18 SNPs under directional positive selection (log10(BF) > 1.5) and Baye-Alpha > 0.97): *rs900414, rs3784515, rs28469755, rs56278789, rs3784519, rs9920839, rs12592414, rs963172, rs3784518, rs3784496, rs7167668, rs12594076, rs2412067, rs34039757, rs4965825, rs747539, rs10400816, rs1871974* and 18 SNPs under balancing selection (log10(BF) > 1.5) and Baye-Alpha < -0.97): *rs11634270, rs11853956, rs11854179, rs11855154, rs12915828, rs2047216, rs28529205, rs34548635, rs36010356, rs4965390, rs4965869, rs55794955, rs55990600, rs59480567, rs62027207, rs6598477, rs71414401, rs7166089*. Both, FDIST and BayeScan found 27 overlapping SNPs under selection processes, of which twelve are under strong positive directional selection (Table 2) and 15 under balancing selection (Table 3). All of these SNPs are intron variants.

Haplotype approaches. In the EHH analysis, all SNPs detected under positive selection by FDIST and BayeScan show a slow decay of the derived extended haplotype compared to the ancestral haplotype in the YRI, CEU, CHB and MXL populations (Supplementary Fig. S1). This decline, however, is comparatively rapid, so that we can assume a very old selective sweep. The iHS scores (Table 4) provided for chromosome 15 show signals higher than one, for *rs1871974* in the YRI and CHB populations, for *rs28469755* and *rs3784496* in the CHB population, for

rs3784515, *rs3784518* and *3784519* in the CHB and MXL population as well as for *rs900414* in the CHB population. Particularly *rs28469755* in the CHB population reveals a value close to two, which can be seen as a sign of a *genome wide iHS signal* (Voight et. al. 2006).

Discussion

The genome-wide analyses of selection processes acting on genetic loci have been attractive indirect strategies to gain insights into human evolution, have implications for human health (Karlsson et al., 2014) and are important for medical advancements (Vallender & Lahn, 2004). This study therefore investigated the natural selection on the PCSK6 gene. The enzyme's functional manifoldness is reflected by its involvement in various metabolic cycles and its participation in numerous signalling pathways. Thereby a great number of substrates is converted by PCSK6 into their active and effective form. On this basis a considerable significance of the proteolytic PCSK6 enzyme in humans can be assumed. This presumption might be confirmed as the PCSK6 selection scan yielded twelve intronic loci under positive and 15 under balancing selection. The discovered plurality of DNA polymorphisms under selection, however, gives us only a vague idea of the evolutionary relevance of this serine endoprotease.

The proposition that balancing and positive selection pressures are acting on PCSK6 is not unreasonable, but remains to be functionally proven. Indeed, a literature research on the 27 polymorphisms under selection detected by LOSITAN and BayeScan reveals, with few exceptions, almost no functional information. However, the SNPs might provide interesting candidate loci for future functional studies.

As twelve loci show high FST values, which is in accordance with the finding of positive selection, a large genetic differentiation in the populations can be assumed (Jobling et al., 2013). These results of strong selective pressures acting on PCSK6, as well as the presence of numerous PCSK6 interaction partners, enable an extensive exploration, interpretation and speculation about the importance of this enzyme. In this discussion therefore a short overview of PCSK6-related mechanisms that might be associated with the detected selection patterns will be presented. Although the literature-based connections made in here depict possible but hypothetic models for the selective pressures acting on PCSK6, they are not claimed to be ultimate and exhaustive. Thus, these theoretic approximations remain unconfirmed without a practical examination and should mainly depict an impetus for further research in this field.

The identification of signals of very recent positive selection can provide information about the adaptation of modern humans to local conditions. The ancestral-susceptibility model (DiRienzo & Hudson, 2005) states in this regard that ancestral alleles, which conferred a selective advantage in earlier human populations, nowadays might increase the risk for diseases under the currently prevailing lifestyle and environment conditions and the advancing life expectancy. The model can be applied to numerous biological traits such as height, endurance, cognition and drug response, which seem to have been subjected to the changing selective pressures during the human evolution (DiRienzo & Hudson, 2005). Similarly, disorders such as near-sightedness, obesity and heart diseases might emerge from changes between the ancient and modern environmental conditions (Keller, 2008). When the

environment changes quickly, mismatches can occur where ancestral alleles are badly adapted to modern lifestyle. Thereby, considering the diversifying environmental context, alleles that were once neutral can be influenced by natural selection and thus become today`s risk alleles (Keller, 2008). Analogous is the susceptibility to several common diseases, mainly shaped by selection processes. Hence, selection analyses can reveal assumptions on the enzyme's involvement in disease-related processes and pathways, which might be representatives of the ancestral-susceptibility model. The positive selection patterns of PCSK6 detected in this study are therefore more closely elaborated concerning its clinical relevance for tumour development, hypertension, osteo- and rheumatoid arthritis as well as pain perception.

Based on the ancestral-susceptibility model, the positive selection acting on the PCSK6 locus might mainly be associated with its function in **tumorigenesis**. When PCSK6 is overexpressed, it increases the susceptibility to carcinogenesis and thus leads to enhanced tumour cell proliferation (Bassi et al., 2010). In addition, the PCSK6 enzyme evidently plays an important role in the development of prostate cancer (D'Anjou et al., 2011), as it is highly expressed in all different clinical stages of prostate tumour tissues. The sustained cancer progression might occur as a result of upregulated or aberrant PCSK6-related processing of growth factors (D'Anjou et al., 2011) such as TGFbeta 1 (Dubois et al., 2001), TGFbeta-related BMPs (Constam & Robertson, 1999), IGF1 (Bassi et al., 2010) or the pro-apoptotic factor TRPS1 (D'Anjou et al., 2011). Furthermore, evidence has implicated the involvement of PCSK6 in skin cancer (Bassi et al., 2010), breast cancer (Cheng et al., 1997) and ovarian cancer (Page et al., 2007). Consequently, the potential role of PCSK6 in the

formation of several tumours and cancer makes the scenario of positive selection feasible. The somatic mutations within tumours that are subject to positive selection might be the basis for the adaptive role of cancer genes during evolution and substantially contribute to transformation, tumour maintenance, expansion and metastasis. Having uncovered the sites affected by positive selection can yield an information gain about PCSK6 in tumorigenesis. In this context, signatures of positive selection delimit those regions of the PCSK6 locus that might be functionally important for the development and progression of various cancer types for further research in the field.

Another function of PCSK6 lies in the production and perception of **pain signals**. A knock-out of PCSK6 in mice leads to a reduced pain sensitivity or protection against pain compared to the wild-type (Malfait et al., 2012). As many phenotypic traits associated with pain probably reflect strong positive selection, the detected positive selection of PCSK6 might be indicative for an assignation of this enzyme an evolutionary-relevant status within nociception. These assumptions would again be in accordance with the ancestral-susceptibility model (DiRienzo & Hudson, 2005). Similarly, the Mas-related genes (MRG) protein was among the first examples of nociception-relation genes to be found under positive selection (Choi & Lahn, 2003). The evolutionary relevance for genes such as MRG or PCSK6 relies on the fact that perception of dangerous stimuli as being painful is crucial to survival. Varying evolutionary, environmental or social frameworks require an altered sensitivity and selectivity to stimuli, making the responsible gene loci (MRG, PCSK6) a substrate of selective pressure and support the ancestral-susceptibility model. Indeed, different human populations show drastically variable responses to identical pain stimulation.

Edwards et al. (2001) for example found statistically significant effects in dependence of ethnic group membership for perceived pain severity and pain-related disability.

PCSK6 is upregulated in the cartilage of patients with **osteoarthritis** (Byun et al., 2010). Its intrinsic SNP *rs900414* between the exons 22 and 23 is significantly associated with symptomatic osteoarthritis in the knee of Caucasians (Malfait et al., 2012). The specific location of this SNP indicates an important role in the variable splicing of the gene (Tsuji et al., 1994). This makes it plausible that positive selection has favoured the functional evolution of this PCSK6 variant. During the search for molecular adaption to selective forces within the PCSK6 locus, we found this polymorphism to be under positive selection by both algorithm programs LOSITAN and BayeScan. Furthermore, Wang et al. (2015) discovered a significantly increased expression of PCSK6 mRNA in the synovial tissues of individuals with rheumatoid arthritis compared to control tissues. Additionally, PCSK6 knock-down experiments reveal a reduced proliferation of rheumatoid arthritis synovial fibroblasts. Based on these observations, an association of PCSK6 and painful joint diseases is probable. These results point to the probability that positive selection on PCSK6 could also be directed towards its function in the development of arthritis.

A haplotype analysis tested the association of PCSK6 and **hypertension** because genetic factors have been estimated to contribute to 20-50% of inter-individual blood pressure variation (Perusse et al., 1991). Thereby Li et al. (2004) discovered a haplotype block within PCSK6 which is linked to diastolic high blood pressure (Li et al., 2004). We also detected a polymorphism within this block as being under

positive selection with the two outlier algorithms used. The SNP *rs1871974* in the PCSK6 gene could therefore be involved in the physiological mechanism of blood pressure regulation and might be again in accordance with the ancestral-susceptibility mode. Although limited data on the role of PCSK6 in blood pressure regulation are available, there may be a potential biological effect of PCSK as a serine protease on the pressure in the vascular system. Two segregation analyses by Xu et al. (1999) have provided evidence for an increased linkage of blood pressure in the genomic region close to the PCSK6 locus (Xu et al., 1999). The carboxy-terminal domain of PCSK6 seems to be essential to effectively process the endothelial lipase and the lipoprotein lipase (Seidah et al., 2008). Hence, these interaction partners might indicate the involvement of PCSK6 in cardiovascular diseases and in the pathological lipid profile. Moreover, Chen et al. (2015) showed that PCSK6 cleaves and stimulates corin, a serine protease that activates natriuretic peptides (Chen et al., 2015). Thereby PCSK6 knock-out mice and mutation studies indicate that PCSK6 functions as a corin activator and is important for sodium homeostasis to regulate blood pressure. This would help explain the positively-selected SNPs discovered by the selection scan performed in this study.

PCSK6 interacts in a wide range of pathways and cellular processes and thus various substrates of this enzyme are known. The enumeration of all PCSK6 targets can also provide an extensive basis to explain the detected balancing selection of 15 gene loci. Balancing selection suggests that the minor alleles bear some fitness disadvantage, which might alter survival or reproduction of an organism. However, it has to be kept in mind that the method of selection detection forbids direct inferences

about a phenotype and conclusions drawn in this discussion are mainly speculative. Therefore SNPs under selection provide mainly viable indicators for future research.

To name but one, PCSK6 plays a key role in activating NODAL and in TGFbeta signalling during the development of **left-right asymmetry** in vertebrates (Constam & Robertson, 2000). The involvement of PCSK6 in axe formation and in the establishment of a left-right asymmetry points to the influence of this locus on laterality and handedness. The genetic association of PCSK6 and handedness has already been demonstrated among persons with dyslexia (Brandler et al., 2013). The current literature and epidemiological data on the distribution of handedness, in which left-handers make up a constant-level minority (10%) in all human populations, would agree with the balancing selection mechanism as detected in PCSK6. More specifically, a negative frequency-dependent selection mechanism with regard to handedness can be assumed (Raymond et al., 1996). Thereby, the less frequent phenotype shows a fitness advantage *only* if it is held constant on a low level. According to Faurie and Raymond (2005), the benefit of left-handers lies in their advantages in fighting situations. The authors provide evidence for this fighting hypothesis in the strong positive correlation they found between homicide rates and left-handedness in traditional societies (Faurie & Raymond, 2005). Thus, negative frequency-dependence might have played an important role in maintaining left-handedness in human populations. Nevertheless, it is still possible that other, still-to-be-determined advantages might be connected with left-handedness.

The function of PCSK6 in the development of lateralization and handedness might be supported by the presence of many SNPs under balancing selection, as detected

in this study. To date, no evidence for the direct association of one of the 15 discovered SNPs and handedness has been found. However, the *rs11855415* polymorphism was revealed in this study to be under balancing selection by a single LOSITAN analysis. This polymorphism is significantly more frequent among right-handed individuals with dyslexia (Scerri et al., 2011).

The EHH approach led to results for a temporal classification by confirming six of the SNPs under directional positive selection (rs1871974, rs784519, rs28469755, rs3784515, rs4965825, rs900414). Considering that EHH only detects relatively recent, respectively incomplete and on-going selective sweeps, those SNPs not verified by EHH may date back longer than 20,000 years (Voight et al., 2006) and could hence be detected only by the two other outlier approaches used.

The field of genomics enables the examination of human adaptation to diseases because natural selection leaves signatures in the genome. Hence, genomic variants might improve survival and reproduction or can have disadvantageous outcomes (Karlsson et al., 2014). The exact nature of the selective pressure on PCSK6 is not completely understood, but this study contributed by detecting and explaining positive and balancing selection patterns on the PCSK6 gene to show a way forward for future research in this field.

Conclusions

Positive as well as balancing selection was identified to act on the proprotein convertase subtilisin/kexin type 6 (PCSK6) gene by applying four different approaches (FDIST, BayeScan, EHH, iHS). These selective signatures can provide

an indirect approach to obtain information of the gene`s relevance. PCSK6 might have a considerable relevance as it is involved in numerous signalling pathways and converts many substrates into their active and effective form. This presumption can be confirmed as the PCSK6 selection scan yielded twelve intronic loci under positive and 15 under balancing selection. Based on this, we provide an overview of PCSK6-related mechanisms potentially associated with the detected selection patterns. Whilst positive selection possibly reflects the role of PCSK6 in tumorigenesis, in nociception, in rheumatoid arthritis as well as in blood pressure regulation, balancing selection might explain the involvement of PCSK6 in vertebral left-right asymmetry and in human handedness.

References

Amigo J, Salas A, Phillips C, Carracedo A. 2008. SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. *BMC Bioinformatics* **9(1):** 428 DOI 10.1186/1471-2105-9-428.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Lukart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9(1):** 323 DOI 10.1186/1471-2105-9-323.

Arning L, Ocklenburg S, Schulz S, Ness V, Gerding WM, Hengstler JG, Falkenstein M, Epplen JT, Güntürkün O, Beste C. 2013. PCSK6 VNTR polymorphism is associated with degree of handedness but not direction of handedness. *PLoS ONE* **8:** e67251 DOI 10.1371/journal.pone.0067251.

Balloux F, Lugon-Moulin N. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**: 155-165 DOI 10.1046/j.0962-1083.2001.01436.x.

Bassi DE, Zhang J, Cenna J, Litwin S, Cukierman E, Klein-Szanto AJ. 2010. Proprotein convertase inhibition results in decreased skin cell proliferation, tumorigenesis, and metastasis. *Neoplasia* 12: 516-526 DOI 10.1593/neo.92030.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of Royal Societies B* 263: 1619-1626 DOI 10.1098/rspb.1996.0237.

Brandler WM, Morris AP, Evans DM, Scerri TS, Kemp JP, Timpson NJ, St Pourcain B, Smith GD, Ring SM, Stein J, Monaco AP, Talcott JB, Fisher SE, Webber C, Paracchini S. 2013. Common variants in left/right asymmetry genes and pathways are associated with relative hand skill. *PLoS Genetics* **9:** e1003751 DOI 10.1371/journal.pgen.1003751.

Byun S, Tortorella MD, Malfait AM, Fok K, Frank EH, Grodzinsky AJ. 2010. Transport and equilibrium uptake of a peptide inhibitor of PACE4 into articular cartilage is dominated by electrostatic interactions. *Archives of Biochemistry and Biophysics* **499:** 32-39 DOI 10.1016/j.abb.2010.04.019.

Chen S, Cao P, Dong N, Peng J, Zhang C, Wang H, Zhou T, Yang J, Zhang Y, Martelli EE, Naga Prasad SV, Miller RE, Malfait AM, Zhou Y, Wu Q. 2015. PCSK6-mediated corin activation is essential for normal blood pressure. *Nature Medicine* **21(9)**:1048-53 DOI 10.1038/nm.3920.

Cheng M, Watson PH, Paterson JA, Seidah N, Chrétien M, Shiu RP. 1997. Pro-protein convertase gene expression in human breast cancer. *International Journal of Cancer* **71:** 966-71 DOI 10.1002/(SICI)1097-0215(19970611)71:6<966::AID-IJC10>3.0.CO;2-4.

Choi SS, Lahn BT. 2003. Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. *Genome Research* **13:** 2252-2259 DOI 10.1101/gr.1431603.

Constam DB, Robertson EJ. 1999. Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases. *Journal of Cell Biology* **144:** 139-149 DOI 10.1083/jcb.144.1.139.

Constam DB, Robertson EJ. 2000. SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes & Development* **14:** 1146-1155 DOI 10.1101/gad.14.9.1146.

D'Anjou F, Routhier S, Perreault JP, Latil A, Bonnel D, Fournier I, Salzet M, Day R. 2011. Molecular Validation of PACE4 as a Target in Prostate Cancer. *Translational Oncology* **4:** 157-172 DOI 10.1593/tlo.10295.

Darwin C, Wallace E. 1958. Proceedings of the meeting of the Linnean Society held on July 1st, 1858. *Journal of the Proceedings of the Linnean Society.* DOI 10.1111/j.1095-8312.1959.tb00854a.x.

Delph, LF, Kelly, JK. 2014. On the importance of balancing selection in plants. *New Phytology* **201(1)**: 45-56 DOI 10.1111/nph.12441.

Diaz FJ, Sugiura K, Eppig JJ. 2008. Regulation of Pcsk6 expression during the preantral to antral follicle transition in mice: opposing roles of FSH and oocytes. *Biology of Reproduction* **78:** 176-83 DOI 10.1095/biolreprod.107.063537.

Di Rienzo A, Hudson RR. 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends in Genetics* **21:** 596-601 DOI 10.1016/j.tig.2005.08.007.

Dubois CM, Blanchette F, Laprise MH, Leduc R, Grondin F, Seidah NG. 2001. Evidence that furin is an authentic transforming growth factor-β1-converting enzyme. *American Journal of Pathology* **158:** 305-316 DOI 10.1016/S0002-9440(10)63970-3.

Edwards CL, Filligim RB, Keefe F. 2001. Race, ethnicity and pain. *Pain* **94:** 133-137 DOI 10.1016/S0304-3959(01)00408-0.

Faurie C, Raymond M. 2005. Handedness, homicide and negative frequency-dependent selection. *Proceedings of Biological Science **272:** 25-28* DOI 10.1098/rspb.2004.2926.

Faurie C, Raymond M. 2013. The fighting hypothesis as an evolutionary explanation for the handedness polymorphism in humans: where are we? *Annals of the New York Academy of Sciences, The Evolution of Human Handedness* **1288**: 110-113 DOI 10.1111/nyas.12159.

Foll M, Gaggiotti OE. 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180:** 977-993 DOI 10.1534/genetics.108.092221.

Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. *Nature Reviews Genetics* **15:** 379-393 DOI 10.1038/nrg3734.

Keller MC. 2008. The Evolutionary Persistence of Genes That Increase Mental Disorders Risk. *Current Direction in Psychological Science* **17 (6)**: 395-399 DOI 10.1111/j.1467-8721.2008.00613.x.

Kiefer MC, Tucker JE, Joh R, Landsberg KE, Saltman D, Barr PJ. 1991. Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15. *DNA Cell Biology* **10:** 757-769 DOI 10.1089/dna.1991.10.757.

Li JP, Wang XB, Chen CZ, Xu X, Hong XM, Xu XP, Gao W, Huo Y. 2004. The association between paired basic amino acid cleaving enzyme 4 gene haplotype and diastolic blood pressure. Chinese Medicine Journal **117:** 382-388.

Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28(2):** 298-299 DOI 10.1093/bioinformatics/btr642.

Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M. 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339(6127):** 1578-1582 DOI 10.1126/science.1234070.

Malfait AM, Seymour AB, Gao F, Tortorella MD, Le Graverand-Gastineau MP, Wood LS, Doherty M, Doherty S, Zhang W, Arden NK, Vaughn FL,

Leaverton PE, Spector TD, Hart DJ, Maciewicz RA, Muir KR, Das R, Sorge RE, Sotocinal SG, Schorscher-Petcu A, Valdes AM, Mogil JS. 2012. A role for PACE4 in osteoarthritis pain: evidence from human genetic association and null mutant phenotype**.** *Annuals of the Rheumatic Diseases* **71:** 1042-1048 DOI 10.1136/annrheumdis-2011-200300.

Mori K, Imamaki A, Nagata K, Yonetomi Y, Kiyokage-Yoshimoto R, Martin TJ, Gillespie MT, Nagahama M, Tsuji A, Matsuda Y.1999. Subtilisin-like proprotein convertases, PACE4 and PC8, as well as furin, are endogenous proalbumin convertases in HepG2 cells. Journal of Biochemistry **125**: 627-33.

Nagahama M, Taniguchi T, Hashimoto E, Imamaki A, Mori K, Tsuji A, Matsuda Y. 1998. Biosynthetic processing and quaternary interactions of proprotein convertase SPC4 (PACE4). *FEBS Letters* **434:** 155-159 DOI 10.1016/S0014-5793(98)00970-3**.**

Nielsen, R. 2005. Molecular signatures of natural *selection. Annual Reviews of Genetics* **39:** 197-218 DOI 10.1146/annurev.genet.39.073003.112420**.**

Nour N, Mayer G, Mort JS, Salvas A, Mbikay M, Morrison CJ, Overall CM, Seidah NG. 2005. The cysteine-rich domain of the secreted proprotein convertases PC5A and PACE4 functions as a cell surface anchor and interacts with tissue inhibitors of metalloproteinases. *Molecular Biology of the Cell* **16:** 5215-5226 DOI 10.1091/mbc.E05-06-0504.

Page RE, Klein-Szanto AJ, Litwin S, Nicolas E, Al-Jumaily R, Alexander P, Godwin AK, Ross EA, Schilder RJ, Bassi DE. 2007. Increased expression of the pro-protein convertase furin predicts decreased survival in ovarian cancer. *Cell Oncology* **29:** 289-299 DOI 10.1155/2007/930321.

Perusse L, Moll PP, Sing CF. 1991. Evidence that a single gene with gender- and age dependent effects influences systolic blood pressure determination in a population based sample. *American Journal of Human Genetics* **49:** 94-105.

Raymond M, Pontier D, Dufour A-B, Moller AP. 1996. Frequency-dependent maintenance of left handedness in humans. *Proceedings of the Royal Society London B* **263:** 1627-1633 DOI 10.1098/rspb.1996.0238.

Raymond M, Pontier D. 2004. Is there geographical variation in human handedness? *Laterality* **9:** 35-52 DOI 10.1080/13576500244000274.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419:** 832-837 DOI 10.1038/nature01140.

Scerri TS, Brandler WM, Paracchini S, Morris AP, Ring SM, Richardson AJ, Talcott JB, Stein J, Monaco AP. 2011. *PCSK6* is associated with handedness in individuals with dyslexia. *Human Molecular Genetics* **20:** 608-614 DOI 10.1093/hmg/ddq475.

Seidah NG, Prat A. 2007. The proprotein convertases are potential targets in the treatment of dyslipidemia. Journal of Molecular Medicine **85: 685-696** DOI 10.1007/s00109-007. 0172-7.

Seidah NG, Mayer G, Zaid A, Rousselet E, Nassoury N, Poirier S, Essalmani R, Prat A. 2008. The activation and physiological functions of the proprotein convertases. International Journal of Biochemistry and Cell Biology **40(6-7)**: 1111-1125 DOI 10.1016/j.biocel.2008.01.030.

Seidah NG, Prat A. 2012. The biology and therapeutic targeting of the proprotein convertases. *Nature Reviews Drug Discovery* **11:** 367-83 DOI 10.1038/nrd3699.

Tortorella MD, Arner EC, Hills R, Gormley J, Fok K, Pegg L, Munie G, Malfait AM. 2005. ADAMTS-4 (aggrecanase-1): N-terminal activation mechanisms. Archives of Biochemistry and Biophysics **444**: 34-44 DOI 10.1016/j.abb.2005.09.018.

Tsuji A, Hine C, Tamai Y, Yonemoto K, Mori K, Yoshida S, Bando M, Sakai E, Mori K, Akamatsu T, Matsuda Y. 1997. Genomic Organization and Alternative Splicing of Human PACE4 (SPC4), Kexin-Like Processing Endoprotease. *The Journal of Biochemistry* **122:** 438-452.

Tsuji A, Hine C, Mori K, Tamai Y, Higashine K, Nagamune H, Matsuda Y. 1994. A novel member, PC7, of the mammalian kexin-like protease family: Homology to PACE4A, its brain specific expression and identification of isoforms. *Biochemical and Biophysical Research Communications* **202:** 1452-1459 DOI 10.1006/bbrc.1994.2094.

Turpeinen H, Ortutay Z, Pesu M. 2013. Genetics of the First Seven Proprotein Convertase Enzymes in Health and Disease. Current Genomics **14:** 453-467 DOI 10.2174/1389202911314050010.

Vallender, EJ, Lahn BT. 2004. Positive selection on the human genome. *Human Molecular Genetics* **13:** R245-R254 DOI 10.1093/hmg/ddh253.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biology* **4:** e72 DOI 10.1371/journal.pbio.0040072.

Wang Y, Wang XH, Fan DX, Zhang Y, Li MQ, Wu HX, Jin LP. 2014. PCSK6 regulated by LH inhibits the apoptosis of human granulosa cells via activin A and TGFβ2. Journal of Endocrinology **222(1)**: 151-160 DOI 10.1530/JOE-13-0592.

Wang F, Wang L, Jiang H, Chang X, Pan J. 2015. Inhibition of PCSK6 May Play a Protective Role in the Development of Rheumatoid Arthritis. *Journal of Rheumatology* **42:** 161-169 DOI 10.3899/jrheum.140435.

Xu X, Rogus JJ, Terwedow HA, Yang J, Wang Z, Chen C, Niu T, Wang B, Xu H, Weiss S, Schork NJ, Fang Z. 1999. An extreme-sib-pair genome scan for genes regulating blood pressure. *American Journal of Human Genetics* **64:** 1694-1701 DOI 10.1086/302405.

Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, Ruff CB, Wang NY, Chakravarti A. 2005. Differential Susceptibility to Hypertension Is Due to Selection during the Out-of-Africa Expansion. *PLoS Genetics* **1**: e82 DOI 10.1371/journal.pgen.0010082.

Author Contributions

S. Kuderer and M. Fieder designed and performed the analysis. S. Kuderer, M. Fieder and S. Kirchengast wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

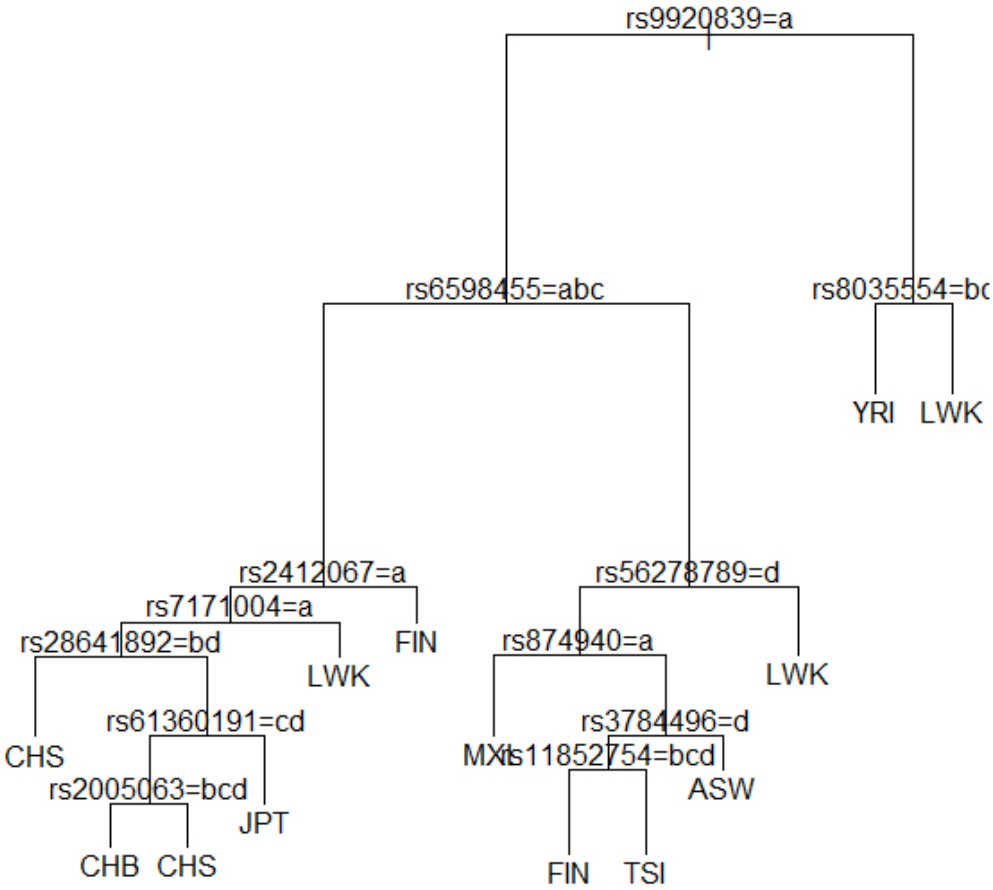
The authors declare no competing financial interests.

683 **Figures**

684 **Figure 1**. Regression tree of SNPs on PCSK6 regressing on 14 populations of the

685 1000 samples.

686



687

688

689

690

691

Figure 2. Detection of outlier SNPs on the PCSK6 gene using LOSITAN. X-axis: estimated heterozygosity (He) values; Y-axis: FST-values. The red area indicates positive selection, the grey area neutrality, and the yellow area balancing selection at P < 0.005. Confidence intervals represent borders between "selection areas". SNPs under selection are displayed as numerical sequences.

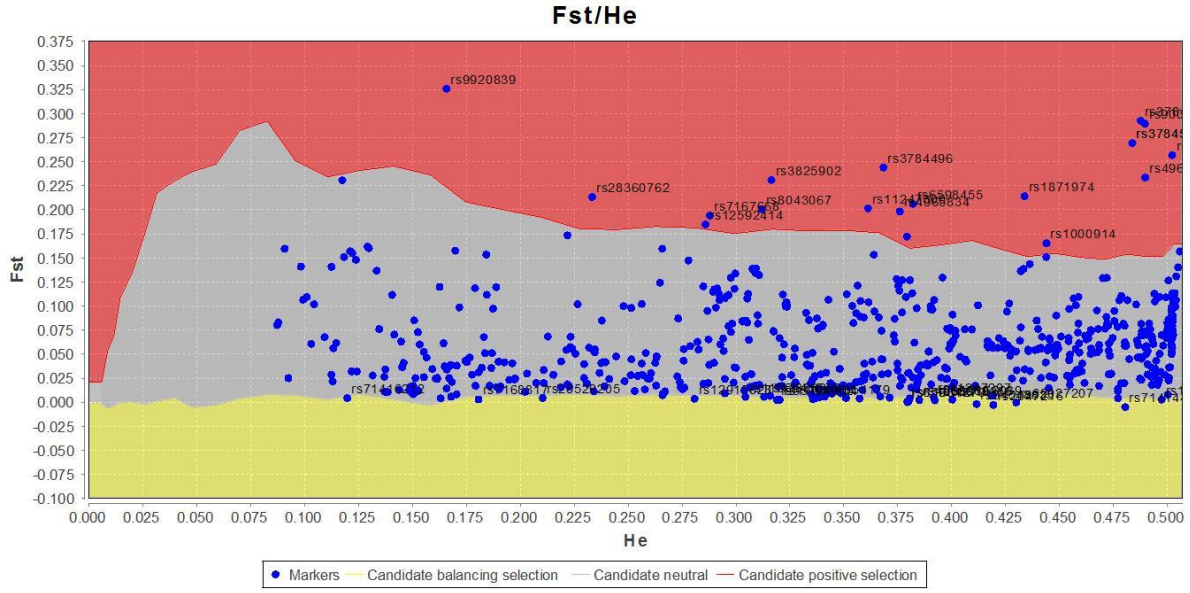
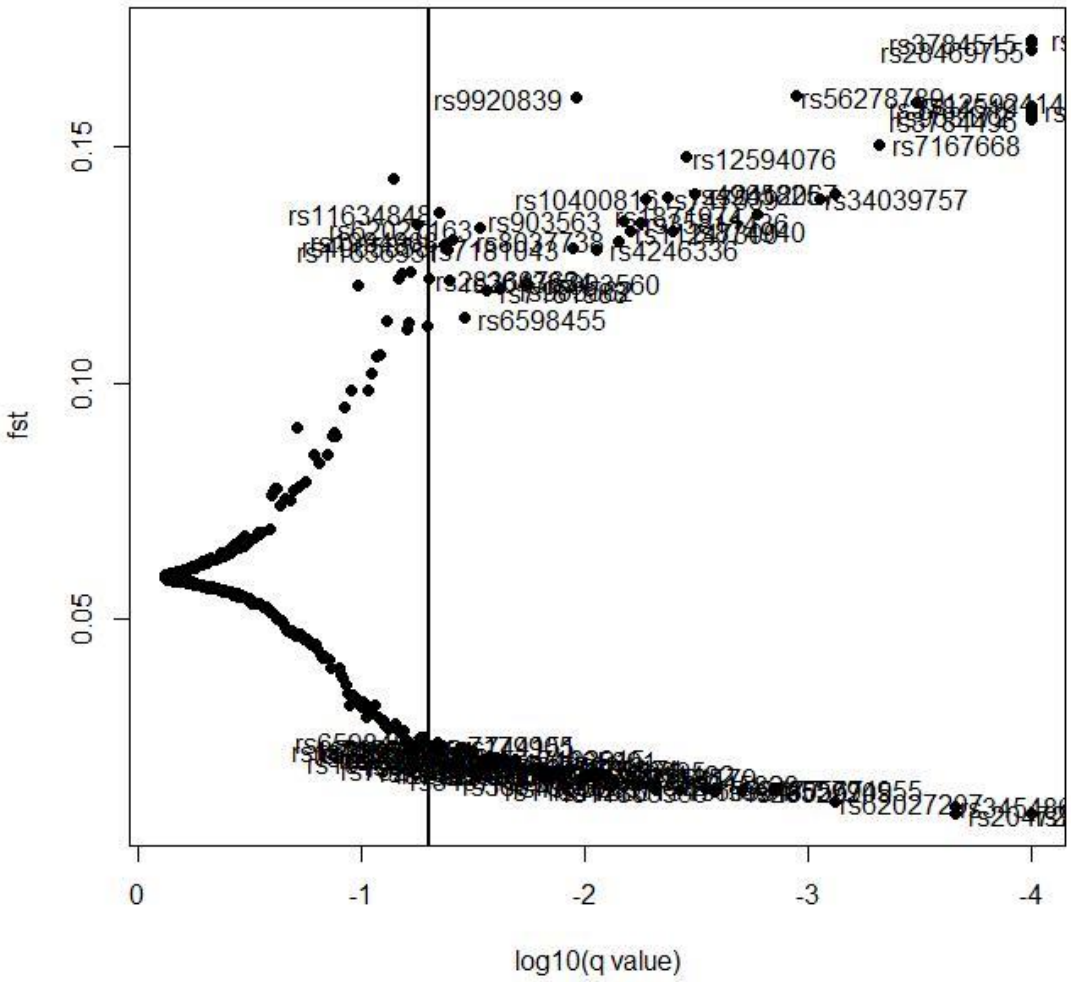


Figure 3. Detection of outlier SNPs in the PCSK6 gene using BayeScan. BayeScan indicates 'very strong' evidence for selection (the vertical line corresponds to log10(BF) > 1.5). log10 (q-value) values are shown on the x-axis and F_{ST}-values on the y-axis.



Tables

Table 1. Populations of the 1000 genome project.

Superpopulation	Population	Number of individuals
Africa (AFR)	African ancestry in Southwest USA (ASW)	61
	Luhya in webuye, Kenya (LWK)	97
	Yoruba in Ibadan, Nigeria (YRI)	88
Europe (EUR)	Utah residents (CEPH) with Northern and Western European ancestry (CEU)	85
	Finnish from Finland (FIN)	93
	British from England and Scotland (GBR)	89
	Iberian populations in Spain (IBS)	14
	Toscani in Italy (TSI)	98
East Asia (ASN)	Han Chinese in Beijing, China (CHB)	97
	Han Chinese South (CHS)	100
	Japanese in Tokyo, Japan (JPT)	89
America (AMR)	Colombians from Medellin, Colombia (CLM)	60
	Mexican ancestry from Los Angeles, USA (MXL)	66
	Puerto Ricans from Puerto Rico (PUR)	55

730 **Table 2**. SNPs identified by both FDIST and BayeScan under strong directional
731 positive selection.

FDIST				BayeScan		
Locus	Het	F_{st} FDIST	P	log10(P0)	Baye-Alpha	F_{st} BayeScan
rs1000914	0.444023	0.16524	0.001301	1.9537	0.98882	0.13865
rs12592414	0.285875	0.184773	0.001913	2.1044	1.0679	0.14779
rs1871974	0.433851	0.214186	0.000122	2.0617	0.97104	0.13649
rs28469755	0.502367	0.256752	0	1000	1.2833	0.17106
rs3784496	0.36837	0.243999	0.000155	1000	1.15	0.15619
rs3784515	0.487685	0.292646	0	1000	1.2912	0.17199
rs3784518	0.483786	0.269345	0	1000	1.1653	0.15734
rs3784519	0.483786	0.269345	0	3.3977	1.178	0.15879
rs4965825	0.489865	0.233589	0	2.442	1.0043	0.13965
rs7167668	0.287938	0.194089	0.001177	2.6187	1.0919	0.15008
rs900414	0.489763	0.289394	0	1000	1.3	0.17307
rs9920839	0.165675	0.325789	0.000445	1.5186	1.1707	0.16156

732 P values for FDIST are reasonably below a significance level of 0.00001. BayeScan
733 log10(PO) values higher than 1.5 as well as Baye-Alpha values above 1 indicate
734 very strong directional positive selection. *Note: F_{st} values between FDIST and*
735 *BayeScan are different due to different F_{st} calculations.*
736
737
738
739
740
741
742
743

744 **Table 3.** SNPs identified by both FDIST and BayeScan under balancing selection.

FDIST				BayeScan		
Locus	Het	P	F_{st} FDIST	log10(PO)	Baye-Alpha	F_{st} BayeScan
rs11634270	0.379966	0.001554	0.000608	1.9155	-1.6983	0.014125
rs11853956	0.313035	0.002338	0.005582	2.0516	-1.7187	0.013611
rs11854179	0.335685	0.001743	0.003056	1.6129	-1.4949	0.016672
rs11855154	0.320336	0.001539	0.0026	2.0417	-1.672	0.014221
rs12915828	0.280766	0.000736	0.00361	1.894	-1.6102	0.015082
rs2047216	0.419421	0.000579	-0.002719	3.0965	-2.2201	0.0088471
rs28529205	0.21031	0.000547	0.004491	2.1518	-1.6856	0.013895
rs34548635	0.411601	0.000703	-0.001854	3.0965	-2.017	0.010491
rs36010356	0.392268	0.000363	0.002007	1.632	-1.6148	0.015373
rs4965869	0.384999	0.00038	0.002138	1.9155	-1.6328	0.014836
rs55794955	0.318958	0.001452	0.002388	2.234	-1.662	0.014161
rs62027207	0.430008	0.000991	-0.000284	2.6187	-1.9322	0.01131
rs6598477	0.379372	0.001456	0.000283	1.9783	-1.7067	0.013972
rs71414401	0.480575	0.000325	-0.005042	3.3977	-2.2046	0.008878
rs7166089	0.400391	0.000469	0.002718	1.7672	-1.5909	0.015381

745 P values for FDIST are reasonably below a significance level of 0.00001. BayeScan
746 log10(PO) values higher than 1.5 as well as Baye-Alpha values below -1 indicate
747 balancing selection.

748

749

750

751

752

753

754 **Table 4**. iHS scores for the YRI, CEU, CHB and MXL population for the SNPs under
755 strong directional positive selection (identified by FDIST and BayeScan).

Locus	iHS YRI	iHS CEU	iHS CHB	iHS MXL
rs1000914	0.605149	0.065135	0.109416	1.04689
rs12592414	0.0183482	NA	0.518972	0.248092
rs1871974	1.05102	0.193538	1.42936	1.03266
rs28469755	0.141949	0.866329	1.97453	1.04179
rs3784496	0.206046	1.30646	0.196389	0.804126
rs3784515	0.242875	0.674884	1.40654	1.048
rs3784518	0.235632	0.668706	1.45598	1.0787
rs3784519	0.235632	0.668706	1.45598	1.14199
rs4965825	0.59076	0.0487176	0.754521	0.959626
rs7167668	0.248535	NA	0.299185	0.2117
rs900414	0.00213491	0.648175	1.62501	0.248092
rs9920839	0.0838233	NA	NA	NA

756 NA indicates that the procedure was not applicable.