Genet. Res., Camb. (2014), vol. 96, e004. © Cambridge University Press 2014
doi:10.1017/S001667231400007X

Novel candidate genes putatively involved in stress fracture predisposition detected by whole-exome sequencing

EITAN FRIEDMAN[1,2]*, DANIEL S. MORAN[3,4], DANNY BEN-AVRAHAM[5], RAN YANOVICH[4] AND GIL ATZMON[5]

[1]*Susanne Levy Gertner Oncogenetics Unit, The Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, 52621 Tel-Hashomer, Israel*
[2]*The Sackler School of Medicine, Tel-Aviv University, Ramat Aviv, Israel*
[3]*Ariel University, Ariel, Samaria, Israel*
[4]*Heller Institute, Sheba Medical Center, 52621 Tel-Hashomer, Israel*
[5]*Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10463, USA*

(*Received 28 December 2013; revised 5 February 2014; accepted 27 February 2014*)

Summary

While genetic factors in all likelihood contribute to stress fracture (SF) pathogenesis, a few studies focusing on candidate genes have previously been reported. The objective of this study is to gain better understanding on the genetic basis of SF in a gene-naive manner. Exome sequence capture followed by massive parallel sequencing of two pooled DNA samples from Israeli combat soldiers was employed: cases with high grade SF and ethnically matched healthy controls. The resulting sequence variants were individually verified using the Sequenom™ platform and the contribution of the genetic alterations was validated in a second cohort of cases and controls. In the discovery set that included DNA pool of cases ($n=34$) and controls ($n=60$), a total of 1174 variants with >600 reads/variant/DNA pool were identified, and 146 (in 127 genes) of these exhibited statistically significant ($P<0{\cdot}05$) different rates between SF cases and controls after multiple comparisons correction. Subsequent validation of these 146 sequence variants individually in a total of 136 SF cases and 127 controls using the Sequenom™ platform validated 20/146 variants. Of these, three missense mutations (rs7426114, rs4073918, rs3752135 in the NEB, SLC6A18 and SIGLEC12 genes, respectively) and three synonymous mutations (rs2071856, rs2515941, rs716745 in the ELFN2, GRK4, LRRC55 genes) displayed significant different rates in SF cases compared with controls. Exome sequencing seemingly unravelled novel candidate genes as involved in SF pathogenesis and predisposition.

1. Introduction

Stress fracture (SF) is a prevalent overuse injury that usually affects bones of the lower extremities in physically active individuals, especially among athletes and military recruits of combat units during basic and advanced training (Hod *et al.*, 2006; Murray *et al.*, 2006) SF is a multifactorial disorder that in all likelihood results from the combined effects of environmental factors in genetically susceptible individuals. Numerous SF risk factors have previously been reported: mechanical [e.g., training intensity and errors (Jones *et al.*, 1993), hip joint anatomy (Giladi *et al.*, 1991), tibia width (Giladi *et al.*, 1987), tibial cortical thickness and femoral neck diameter (Cosman *et al.*, 2013)]; environmental [e.g., shoes (Schwellnus *et al.*, 1990) and training surface (Albisetti *et al.*, 2010)]; and behavioural [e.g., nutritional habits (Frusztajer *et al.*, 1990), smoking (Altarac *et al.*, 2000), motivation (Hallel *et al.*, 1976) and pre-military induction training intensity (Cosman *et al.*, 2013)]. In the largest study published to date encompassing 583 681 US military recruits over a 10-year period, the most significant SF risk factors were older age, lower body weight, lower body mass index (BMI) and race other than African American (Knapik *et al.*, 2012). A large study that

* Corresponding author: Head, the Susanne Levy Gertner Oncogenetics Unit, The Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, 52621 Tel-Hashomer, Israel. Tel.: 972 3 530 3173. Fax: 972 3 535 7308. E-mail: eitan.friedman@sheba.health.gov.il; eitan211@netvision.net.il

included data of 311 pelvic, femoral and knee SF in Finnish military recruits that included 102 515 person years, reported that advanced age and female gender were the most significant SF risk factors in Finnish military personnel (Mattila *et al.*, 2007).

Multiple lines of evidence imply that genetic predisposition may contribute to SF: SF in monozygotic twins (Singer *et al.*, 1990), in the paediatric age group (Bachmann *et al.*, 2011; Lee *et al.*, 2005), the high rate (10·6%) of SF recurrence within 1 year at different anatomical sites (Giladi *et al.*, 1986) and familial clustering of SF or bone diseases in American military personnel (Friedl *et al.*, 1992) and Israeli soldiers with SF (Givon *et al.*, 2000). However, the genetic basis of SF and the precise genes involved in this disorder are unknown.

There are a few studies reporting limited genetic analyses of sequence variants in candidate genes in a small number of SF cases and controls. Valimaki *et al.* (2005) genotyped 15 Finnish soldiers with SF and compared the rates of two polymorphisms in the oestrogen receptor alpha gene – *ESR1* (MIM 133430), *Xba*I (351A→G) and *Pvu*II (397T→C), and to the trinucleotide CAG repeats in the androgen receptor gene – *AR* (MIM 313700) in these 15 cases with that of 164 non-SF controls. No differences in polymorphisms rates were noted between cases and controls. Chatzipapas *et al.* (2009) assessed the rates of four polymorphisms (FokI, BsmI, ApaI and TaqI) in the vitamin D receptor– *VDR* (MIM 601769) in 32 Greek individuals with SF and 32 controls, and reported a significant 2·7-fold increased SF risk associated with the FokI allele. Korvala *et al.* (2010) analysed single nucleotide polymorphism (SNPs) in eight candidate genes in 72 Finnish soldiers with femoral neck SF and reported that a specific VDR haplotype and a SNP in the calcium-sensing receptor (CaSR MIM 601199) gene were more prevalent among cases than controls (OR = 3·22, 95% CI 1·38–7·49), especially in individuals with low BMI. We have previously reported on the association between the length of the CAG repeat in the androgen receptor gene and SF risk in Israeli soldiers (Yanovich *et al.*, 2011), and statistically insignificant associations between several SNPs located within candidate genes and SF in Israeli soldiers (Yanovich *et al.*, 2012). Cosman *et al.* (2013) reported no significant differences in the rates of SNPs and polymorphisms between US military cadets with SF ($n = 43$ males; $n = 26$ females) compared with 712 male controls and 110 female controls in the VDR, COL1A1 (MIM 120150) and estrogen receptor (ER) genes. Thus, the aim of the study was to assess the association of intragenic sequence variants in a gene-naive manner, using whole-exome capture and massive parallel sequencing in Israeli Defense Forces (IDF) combat soldiers using a two-tires discovery-validation, case–control study design.

2. Materials and methods

(i) *Participant identification and recruitment*

Active duty soldiers in the IDF, who were referred to the central orthopaedic clinic of the IDF for clinical examination and whole-body bone scan for symptoms compatible with SFs between 1 January 2005 and 31 December 2010, were eligible for participation. Controls were classified as having no evidence of SF after both clinical examination and bone scan, and cases were soldiers with grade 3–4 SF or multiple grade 2 SF. This study was approved by the IDF Ethics Committee and the Ministry of Health's Ethics Committee for genetic studies, according to the ethical principles for medical research involving human subjects. Each participant, after signing a written informed consent, filled a questionnaire that detailed demographic data, personal and family history of SF and bone diseases, engagement in sports, smoking and alcohol consumption habits, and additional data relevant to SF risk, as previously detailed (Yanovich *et al.*, 2012).

(ii) *Clinical and imaging evaluation*

All participants were clinically evaluated by an orthopaedic surgeon whose physical examination focused on the lower limbs. Imaging included Technetium-99 m methylene diphosphonate (Tc^{99}) whole-body bone scan as routinely practiced in the Institute of Nuclear Medicine at the Medical Services and Supply Center for IDF soldiers (Moran *et al.*, 2008). The results of the bone scan were interpreted by two different expert radiologists, and based on the results of the bone scan, the soldiers were classified as either having no evidence of SF, or having grade 1–4 SF, according to the practiced criteria and protocols (Zwas *et al.*, 1987). Individuals with a single grade 1 or 2 SF as well as individuals with metatarsal SF were excluded from the study.

(iii) *DNA extraction*

Blood (20 ml) was drawn into EDTA-containing tubes by certified phlebotomists and the DNA extraction was completed within 4 h from the blood sample collection. DNA was extracted from peripheral blood leucocytes using the PUREGENE kit by Gentra Systems Inc. (Minneapolis, MN), following the manufacturer's recommended protocol.

(iv) *DNA pools*

Two pooled DNA samples from Israeli soldiers were generated: cases were soldiers with high grade SF ($n = 34$) and controls were ethnically and age-matched soldiers with no evidence of SF ($n = 60$). Equimolar

amounts of DNA from each of the 34 cases for the 1st DNA pool and each of the 60 for the controls DNA pool were prepared according to the method previously described for whole-exome sequencing (Flanagan *et al.*, 2013). Each of the pools was run on 2% agarose to assess the quality of the genomic DNA; degraded samples were removed and replaced by a non-degraded sample.

(v) *Genetic analyses outline*

(a) *Discovery phase genotyping*

The two DNA pools, the cases ($n = 34$) and the controls ($n = 60$) were initially subjected to whole-exome capture and massive parallel sequencing (see below).

(b) *Validation phase*

For the initial validation of the SNPs emerging from the pooled analysis, 32 SF subjects and 55 controls, all from the initial pooled genotyped participants, were subsequently individually genotyped using the Sequenom™ assay (see below). As an additional validation, 104 SF cases and 72 controls, who were similarly identified and recruited but were not genotyped in the Whole exome sequencing (WES) phase, were individually genotyped for the same sequence variants that emerged as significant from the pooled WES phase by applying the Sequenom™ assay. The flow of genotyping and filtering is shown in Fig. 1.

(vi) *Capture array*

NimbleGen 2.1M Human Exome v1.0 Sequence Capture array (Roche NimbleGen, Inc. Madison, WI) was used to target the entire human exome for capturing.

(vii) *Next generation sequencing*

The capture sequences were subsequently subjected to massive parallel sequencing using the HiSeq2000 apparatus (Illumina, San Diego, CA).

(viii) *Sequenom™ genotyping technology*

For validation of a selected number of SNPs for each of the study participants, the Sequenom™ technology was applied, using the primer design and methodology as previously described (Jurinke *et al.*, 2002).

(ix) *Haplotype analysis*

The haplotype association test between cases and controls was performed using Haploview 4.2 (http://www.broadinstitute.org/scientificcommunity/science/programs/medical-and-population-genetics/haploview/haploview).

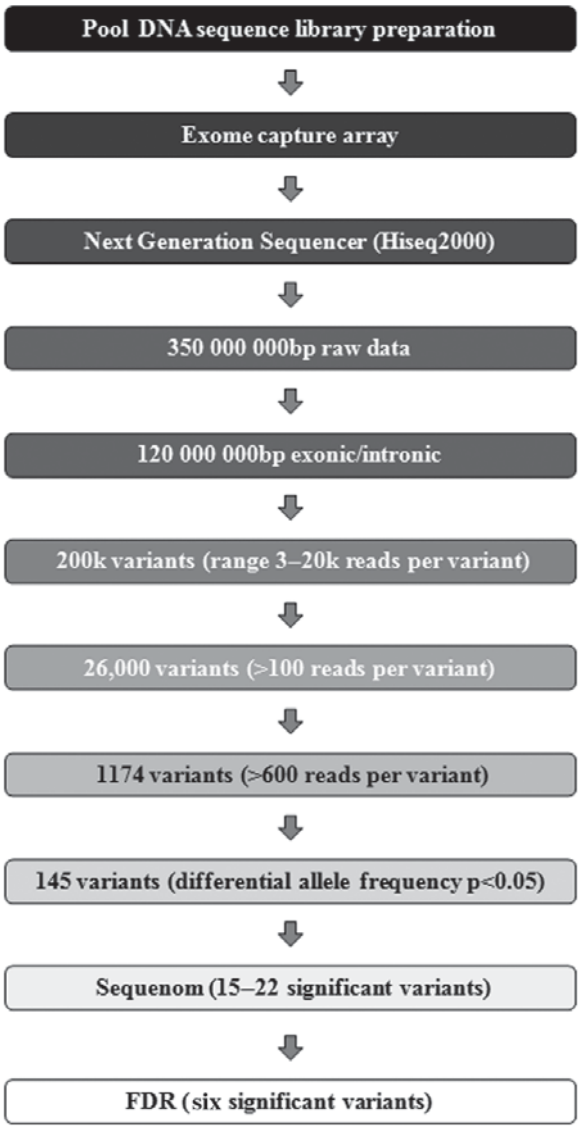

Fig. 1. Schematic view of the pipeline analysis.

(x) *Ingenuity pathway analysis*

We implemented the ingenuity pathway analysis (IPA) (www.ingenuity.com) to identify and classify networks, pathways, biological processes and molecular functions of the genes that emerged from the current study as putatively involved in SF predisposition.

(xi) *Statistical analysis*

Analyses were carried out by 'Wiki-based Automated Sequence Processor (WASP)' (Dubin *et al.*, 2010), of sequence data first generate descriptive statistics to assess quality. Sequence variants that passed the QC threshold, were then aligned to the human reference genome using the Burrows–Wheeler alignment (BWA; Li & Durbin, 2009; Li & Durbin, 2010), Polymerase Chain Reaction (PCR) duplicates eliminated with Picard MarkDuplicates (http://picard.sourceforge.net)

and local re-alignment of indel regions achieved with Genome Analysis Toolkit (GATK; McKenna *et al.*, 2010; DePristo *et al.*, 2011). Determination of population frequency (HapMap3 and 1000 genomes) and prediction of functional impact by BLOSUM62, Polyphen2 and SIFT (Sunyaev *et al.*, 2001; Ng & Henikoff, 2002; Eddy, 2004; Adzhubei *et al.*, 2010) were implemented as part of the analysis pipeline. The mpileup option within Samtools (align sequences from multiple files) (Li *et al.*, 2009) and custom scripts that were developed at the Albert Einstein College of Medicine (Bronx, NY) were used to determine the average read depth per exon.

The raw data from the sequencing pipeline were uploaded to the GeneSifter engine (www.geospiza.com) where a variants list was generated, and chromosome, position, dbSNP, variant, type change, description, read depth, score, gene and function were provided. A list of candidate sequence variants based on the above outlined scheme was generated (Fig. 1).

Case–control analysis was performed using JMP Genomics 6.0 and probability of genotype, allele, dominant and recessive was calculated. False discovery rate (FDR) (*Q* value of the R package) adjustment for multiple comparisons was implemented at the initial screen analysis followed by the chi-square test of the allele distribution in the DNA pool. Finally, we applied the algorithm suggested by Mirina *et al.* (2012) to correct for any potential bias that stem from the gene size, for all variants that have had at least 600× coverage.

3. Results

(i) *Participants' characteristics*

Overall, 161 male soldiers age ranging between 18 and 23 years (mean 20·1 ± 0·9 years), with grade 3–4 SF or multiple grade 2 SF were assigned 'case' status, and 147 male soldiers age ranging between 18 and 25 years (mean 20·1 ± 1·1 years), were classified as having no evidence of SF and were assigned 'control' status. Most cases (83%) and controls (77%) were within 4 months of recruitment and initiation of their basic military training. There were no statistically significant differences in anthropometric measures, physical activity habits prior to enrolment, smoking habits and alcohol consumption between the two study groups, as previously detailed (Yanovich *et al.*, 2012). The ethnic origin of the study population consisted of 45·1% (139/308) Ashkenazi, 41·2% (127/308) non-Ashkenazi, and in 13·6% (42/308) ethnicity could not be ascertained as a result of refusal of the participants to answer this specific question. In the SF group, 105 participants (65·2%) were diagnosed with multiple SFs: 52·3% (55/105) suffered from bilateral tibial fractures, and 15·2% (16/105) had bilateral femoral fractures. Among the participants who sustained single SF ($n = 54$), 25 (46·2%) had tibial SF, 22 (40·7%) had femoral SF, and in seven (12·9%) SF was located at the fibula.

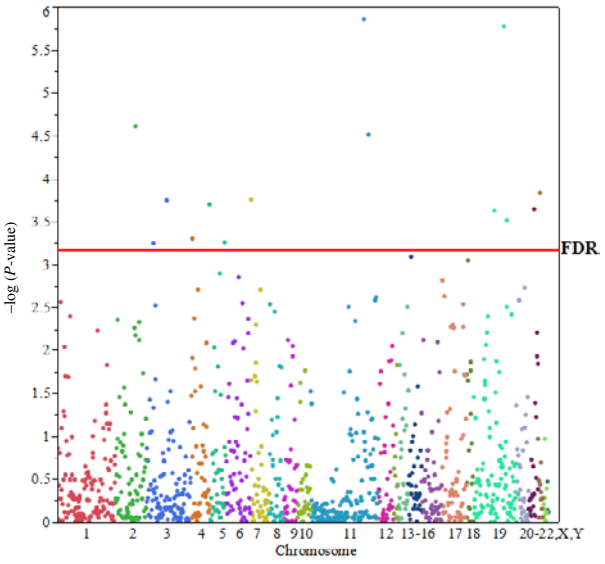

Fig. 2. Manhattan plot shows that functional sequence variants are distributed across all chromosomes in the pooled groups. Results are plotted as negative log10 transformed *P*-values from a genotypic association chi test using allele frequency differences. The red line highlighted the FDR significant threshold.

(ii) *Genetic analyses – genotyping data*

(a) *Discovery phase genotyping*

In the initial screen, 202 406 and 144 217 sequence variants (case and control pools, respectively) were generated. Following initial QC filtering, a list of 67 408 variants with 3–20 000 reads/variant was generated. Of these, a list of 1174 variants with more than 600 reads/variant (allowing 10× coverage for each individual in the largest pool) in both DNA pools were used for comparison between cases and controls. Allele frequencies in each group (cases and controls) during this phase were corrected for the number of alleles in each pool divided by the pool size (i.e., number of individuals in each pool). In total, 146 sequence variants showed statistically significant ($P < 0·05$) differences in rates between cases and controls. Correction for FDR of these 1174 variants of the initial screened pools, revealed 14 sequence variants that had maintained a statistically significant differential rates between cases and controls (Figure 2, Supplemental Table 1). In order to be as inclusive as possible, the list of 146 candidate sequence variants (including the 14 that survived the FDR correction) was used for designing 11 multiplexes for

Sequenom™ assays (Supplemental Table 1) that were further used for individual genotyping as a first phase validation of subjects from the initial screened DNA pools.

(b) *Initial validation*

This step of validation of the initial case–control analysis resulted in 15 sequence variants in 15 genes that displayed either genotypic or allelic nominal significant ($P<0{\cdot}05$) different rates between cases and controls. Five of these 15 sequence variants in five genes (rs7426114, rs4073918, rs3801369, rs716745 and rs3752135) maintained statistical significance after FDR correction (Supplemental Table 2).

(c) *Independent validation phase*

Twelve sequence variants of the 146 initial screened were validated in the second independent group of cases and controls, and two variants in two genes, maintained statistical significance after FDR correction (Supplemental Table 3).

(d) *Combined analysis step*

The two genotyping steps of the 146 candidate sequence variants were combined and re-analysed. Following this analysis, 20 sequence variants in 18 genes (Table 1) displayed a nominal significant ($P<0{\cdot}05$) threshold in one of the models used for analysis (additive, allele, dominant and recessive). Of these, six variants in six genes maintained significance after FDR correction (rs7426114, rs2515941, rs4073918, rs716745, rs3752135 and rs2071856; Table 1). Notably, one sequence variant (rs4073918) in the *SLC6A18* gene consistently displayed significantly different rates between SF cases and their controls in all stages of analysis (initial pool, each of the validation sets and the combined analysis steps) and significance was maintained after FDR correction. In addition, two genes (CR1 and OR10H3) harboured two candidate sequence variants each (rs3811381, rs2296160 and rs11670007, rs11670326, respectively) that were validated in the combined analysis step (Table 1). Subsequent haplotype analysis showed that the GG haplotype in the CR1 gene was over-represented among cases compared with controls ($P=0{\cdot}008$) (Supplement Table 4).

(e) *Pathway analysis*

Employing IPA (http://www.ingenuity.com) for the six candidate genes that maintained significance in the combined analysis after FDR (NEB, GRK4, SLC6A18, LRRC55, SIGLEC12 and ELFN2) several major pathways emerged: cell development, morphology, survival and death, cell-to-cell signalling and interaction, humoral immune response, inflammatory response, nervous system development and function, and tissue development (Fig. 3*a*). Functional (canonical) analysis of these candidate genes was done; they are primarily involved in cell communication, physiology and proliferation (Fig. 3*b*). Although SLC6A18 was the only gene harbouring a sequence variant that consistently showed different rates at all analyses stages, it did not classify to within any of the above-mentioned networks and pathways. IPA analysis highlighted the role played by the UBC gene in pathway associations with the above-listed genes (NEB, GRK4, ELFN2 and CR1). Noteworthy, another algorithm [The UniProt Gene Ontology (GO) annotation program (http://www.ebi.ac.uk/GOA/)] failed to reproduce these network and pathway associations. However, the GO analysis highlighted the role that the NEB gene may play in determining SF phenotype based on its function, process and pathway involvement (Supplemental Table 5). While IPA identifies the functional networks and physiologic processes resulting while highlighting the gene–gene interaction, the GO analysis focuses on the gene product properties based on cross-species functional studies.

4. Discussion

The present study identified several novel candidate genes that are seemingly involved in SF predisposition and pathogenesis: GRK4 (OMIM# 137026), a guanine nucleotide-binding protein whose function has been associated with hypertension (Allayee *et al.*, 2001) and activation of kidney dopamine receptor signalling cascades in human proximal tubule cells (Villar *et al.*, 2009). Nebulin (NEB-OMIM#161650), a muscle protein of the cytoskeletal matrix; SLC6A18 (OMIM# 610300), specific transporter for neurotransmitters; LRRC55 (OMIM# 615213), a protein that regulates voltage- and calcium-activated potassium channels involved in contractile tone of smooth muscles (Yan & Aldrich, 2012). SIGLEC12 (OMIM# 606094) a putative adhesion molecule that mediates osteoclast differentiation (Takahata *et al.*, 2007) and ELFN2 (UniProtKB: Q5R3F8) whose function is to inhibit phosphatase activity of protein phosphatase 1 (PP1), a complex that enhances bone accumulation (Zhang *et al.*, 2009). Pathway analysis of these genes revealed that they are primarily involved in cell development, morphology, survival and death, cell-to-cell signalling and interaction, humoral immune response, inflammatory response, nervous system development and function, and tissue development. Furthermore, pathway analyses highlighted the role that the NEB and the UBC genes

Table 1. *Significant variants*

CHR	Pos	dbSNP	Gene	Base conversion	Type	AA Conversion	P Genotype	P Allele	P Dominant	P Recessive	MA	MAF Case	MAF Control
1	207790088	rs3811381	CR1	C↔S	M	*P*↔R	0·02	0·04	0·02	1·000	G	0·07	0·17
1	207795320	rs2296160	CR1	A↔R	M	T↔A	0·05	0·01	0·09	0·658	A	0·12	0·07
2	152527572	rs7426114	NEB	C↔Y	M	V↔M	0·02	0·05	0·01	0·733	C	0·26	0·19
2	231077112	rs28930679	SP110	G↔R	M	A↔V	0·02	0·10	0·01	0·005	A	0·27	0·33
3	52555957	rs9853056	STAB1	T↔Y	S	–	0·09	0·11	0·03	0·067	T	0·48	0·4
4	946226	rs11552301	TMEM175	T↔Y	S	–	0·05	0·97	0·31	0·038	T	0·32	0·32
4	2993980	rs2515941	GRK4	T↔Y	S	–	0·02	0·98	0·23	0·049	C	0·31	0·29
4	119736796	rs2389688	SEC24D	A↔M	S	–	0·01	0·01	0·14	0·004	C	0·4	0·53
5	1244425	rs4073918	SLC6A18	C↔Y	M	*P*↔L	0·02	0·43	0·68	0·016	C	0·27	0·24
5	150578574	rs3734038	CCDC69	A↔R	S	–	0·04	0·30	0·07	0·160	G	0·25	0·21
10	106907440	rs41426648	SORCS3	C↔Y	S	–	0·03	0·51	0·05	0·010	C	0·49	0·46
11	34969112	rs11539202	PDHX	A↔R	M	T↔A	0·03	0·02	0·14	0·900	G	0·17	0·26
11	56954846	rs716745	LRRC55	C↔Y	S	–	0·12	0·03	0·04	0·196	T	0·43	0·3
12	80014907	rs78187003	PAWR	A↔W	S	–	0·06	0·01	0·16	0·751	A	0·16	0·27
14	21623290	rs7145814	OR5AU1	T↔Y	M	I↔V	0·18	0·05	0·11	0·473	T	0·29	0·2
15	72040774	rs1872056	THSD4	C↔Y	S	–	0·14	0·07	0·05	0·068	T	0·16	0·24
19	15852363	rs11670007	OR10H3	G↔R	M	R↔H	0·15	0·18	0·07	0·047	A	0·31	0·17
19	15852544	rs11670326	OR10H3	C↔Y	S	–	0·10	0·03	0·09	0·402	T	0·2	0·17
19	52000624	rs3752135	SIGLEC12	T↔G	M	Y↔S	0·05	0·19	0·57	0·706	T	0·14	0·19
22	37770630	rs2071856	ELFN2	G↔R	S	–	0·12	0·05	0·14	0·648	A	0·32	0·21

M, missense; S, synonymous; MA, minor allele; MAF, minor allele frequency.

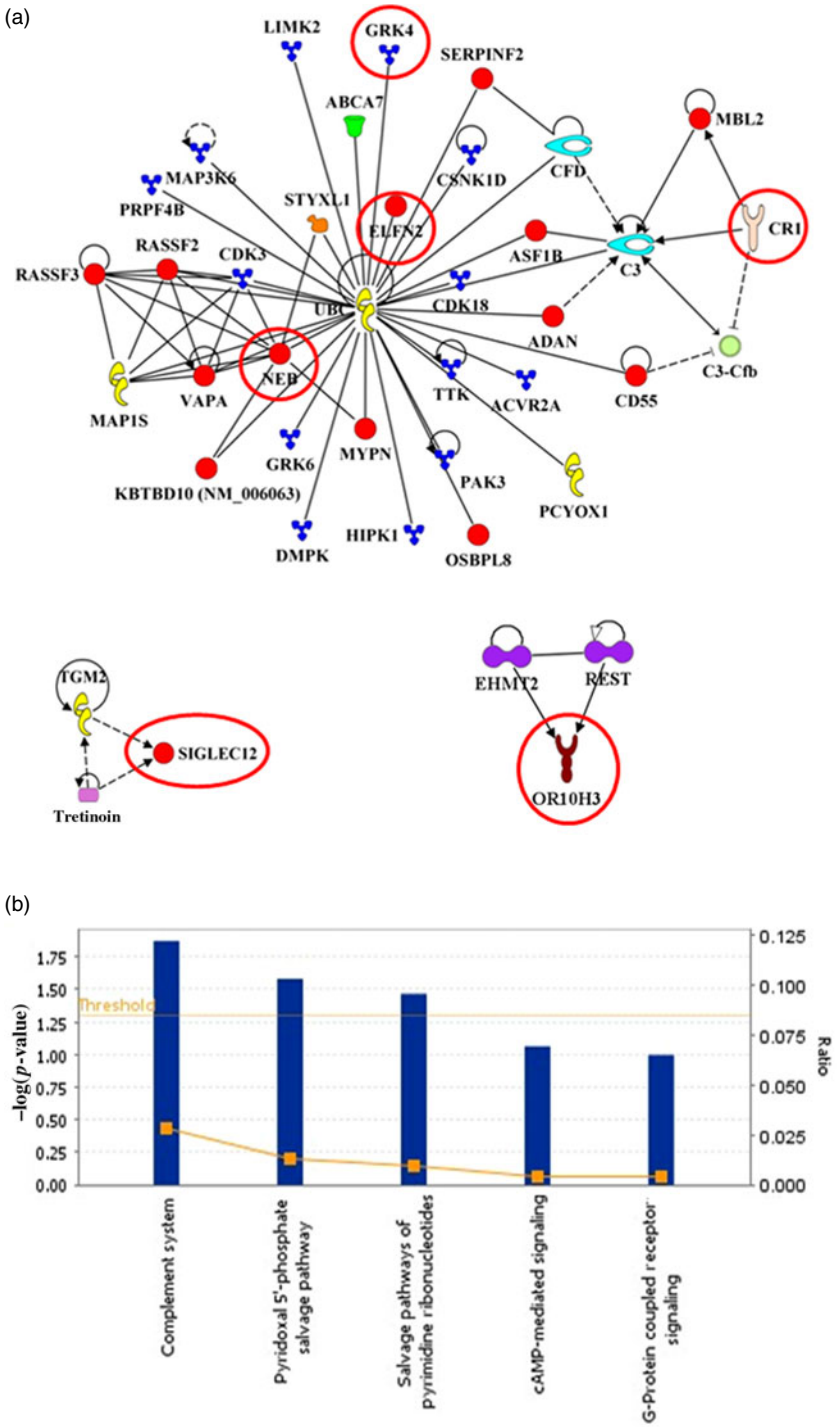

Fig. 3. Computational analysis of the genes that differed significantly among cases (SF) and controls, using two complimentary approaches: (*a*) IPA Network analysis. Genes are represented as nodes; solid and hatched lines depict direct and indirect interactions, respectively. Human gene functions are colour-coded as follows: Red, unknown; Blue, kinase; Yellow, enzyme; Beige, transmembrane receptor; Light blue, peptidase; Green, transporter; Orange, mphosphatase, Light green, complex; Magenta, chemical endogenous mammalian; Purple, transcription regulator; Brown, G-protein-coupled receptor; and (*b*) functional (canonical) analysis of candidate genes, ordered according to the Fisher's test.

may play in SF pathogenesis. While the NEB gene may plausibly and mechanistically be associated with SF pathogenesis, given its role in muscle cytoskeleton matrix, the UBC gene (MIM # 191340) that encodes for a ubiquitin is associated with cell–cell adhesion, DNA damage repair, intracellular degradation and other processes, may be less obvious in its association with SF pathogenesis. Only subsequent

validation of the contribution of these genes and pathways in additional SF cases as well as functional assessment of their contribution to the phenotype in animal model may unravel the mechanisms by which these genes and pathways may lead to SF. Noteworthy, some of these genes and pathways have not been considered 'classical SF candidate genes' and have never been reported to be analysed as candidate SF genes. The few previous studies on the genetic predisposition to SFs were all focused on seemingly logical candidate genes: genes involved in bone formation and bone-associated disorders. Despite several reports on a significant statistical association between SF phenotype and a genetic variant [e.g., CAG repeat in the androgen receptor gene (Valimaki *et al.*, 2005), VDR polymorphisms (Chatzipapas *et al.*, 2009)] all studies were notably based on a limited number of cases and have analysed a rather limited number of sequence variants and genes. The involvement of genes along the novel pathways that are indicated by this preliminary study may shed new light on the pathogenesis of SF. The results of this study are consistent with the notion that the genes directly involved in bone formation, metabolic and inherited bone diseases and osteoporosis contribute little, if anything, to SF pathogenesis. In support of this notion, analysis of the chromosomal location of the sequence variants found in the present study as putatively associated with SF pathogenesis ($n = 146$) showed that none co-localized with any of the SNPs or genes previously associated with metacarpal bone geometry (Karasik *et al.*, 2008), trabecular and cortical volumetric bone mineral densities (Paternoster *et al.*, 2010, 2013), osteoporosis-related phenotype (Karasik *et al.*, 2012), bone mass, hip geometry and fractures (Boudin *et al.*, 2013; Garcia-Ibarbia *et al.*, 2013), bone mineral density, cortical bone thickness and bone strength (Zheng *et al.*, 2012), bone mineral density loci and osteoporotic fracture (Duncan *et al.*, 2011; Estrada *et al.*, 2012), and femoral neck bone geometry (Zhao *et al.*, 2010) (Supplemental Table 6), including the ones that were reviewed by Ralston & Uitterlinden (2010) (Supplemental Table 7).

If the results of the present study are validated in a larger study, of ethnically diverse individuals with SF, the data can be used for several clinical applications: identifying pre-symptomatic individuals at high risk for developing SF and offering them a modified training regimen, and possibly using the up-regulated gene products as surrogate markers for diagnosing SF without using radiation-emitting modalities.

This study has several limitations that should be pointed out. It focused on Jewish Israeli active duty soldiers, and the validation of the findings in other ethnically diverse population is further stressed. Although this is the largest genetic study pertaining to SFs, the number of analysed individuals is still limited. While DNA pooling is an efficient mean of lowering the costs of whole-exome sequencing, unequal representation of all participants in the DNA pool and skewed DNA amplification may have eliminated some additional candidate regions and genes. Lastly, the molecular mechanisms leading to SF in soldiers during basic training may differ than those underlying the same phenotype in elite training athletes.

In conclusion, using whole-exome capture and sequencing, novel candidate genes and pathways have emerged as putatively contributing to SF pathogenesis and predisposition.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S001667231400007X.

This study was funded by a grant from TATRC (W81XWH-09-2-0054) to Eitan Friedman and Daniel S. Moran.

Conflict of interest statement

All authors declare that they have no conflict of interest regarding the data published.

References

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

Albisetti, W., Perugia, D., De Bartolomeo, O., Tagliabue, L., Camerucci, E. & Calori, G. M. (2010). Stress fractures of the base of the metatarsal bones in young trainee ballet dancers. *International Orthopaedics* **34**, 51–55.

Allayee, H., de Bruin, T. W., Dominguez, M. K., Cheng, L. S., Ipp, E., Cantor, R. M., Krass, K. L., Keulen, E. T., Aouizerat, B. E., Lusis, A. J. & Rotter, J. I. (2001). Genome scan for blood pressure in Dutch dyslipidemic families reveals linkage to a locus on chromosome 4p. *Hypertension* **38**, 773–778.

Altarac, M., Gardner, J. W., Popovich, R. M., Potter, R., Knapik, J. J. & Jones, B. H. (2000). Cigarette smoking and exercise-related injuries among young men and women. *American Journal of Preventive Medicine* **18**, 96–102.

Bachmann, M., Gaston, M. S. & Hefti, F. (2011). Supracondylar stress fracture of the femur in a child. *Journal of Pediatric Orthopaedics B* **20**, 70–73.

Boudin, E., Steenackers, E., de Freitas, F., Nielsen, T. L., Andersen, M., Brixen, K., Van Hul, W. & Piters, E. (2013). A common LRP4 haplotype is associated with bone mineral density and hip geometry in men – data from the Odense androgen study (OAS). *Bone* **53**, 414–420.

Chatzipapas, C., Boikos, S., Drosos, G. I., Kazakos, K., Tripsianis, G., Serbis, A., Stergiopoulos, S., Tilkeridis, C., Verettas, D. A. & Stratakis, C. A. (2009). Polymorphisms of the vitamin D receptor gene and stress fractures. *Hormone and Metabolic Research* **41**, 635–640.

Cosman, F., Ruffing, J., Zion, M., Uhorchak, J., Ralston, S., Tendy, S., McGuigan, F. E., Lindsay, R. & Nieves, J. (2013). Determinants of stress fracture risk in United States Military Academy cadets. *Bone* **55**, 359–366.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–498.

Dubin, R., Jing, Q., O'Broin, P., Calder, B., McLellan, A., Moskowitz, D., Suzuki, M., & Greally, J. M. (2010). WASP: Wiki-based automated sequence processor for epigenomics and genomics applications. *Journal of Biomolecular Techniques* **21**(3 Suppl), S11.

Duncan, E. L., Danoy, P., Kemp, J. P., Leo, P. J., McCloskey, E., Nicholson, G. C., Eastell, R., Prince, R. L., Eisman, J. A., Jones, G., Sambrook, P. N., Reid, I. R., Dennison, E. M., Wark, J., Richards, J. B., Uitterlinden, A. G., Spector, T. D., Esapa, C., Cox, R. D., Brown, S. D., Thakker, R. V., Addison, K. A., Bradbury, L. A., Center, J. R., Cooper, C., Cremin, C., Estrada, K., Felsenberg, D., Gluer, C. C., Hadler, J., Henry, M. J., Hofman, A., Kotowicz, M. A., Makovey, J., Nguyen, S. C., Nguyen, T. V., Pasco, J. A., Pryce, K., Reid, D. M., Rivadeneira, F., Roux, C., Stefansson, K., Styrkarsdottir, U., Thorleifsson, G., Tichawangana, R., Evans, D. M., & Brown, M. A. (2011). Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genetics* **7**, e1001372.

Eddy, S. R. (2004). Where did the BLOSUM62 alignment score matrix come from? *Nature Biotechnology* **22**, 1035–1036.

Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y. H., Duncan, E. L., Ntzani, E. E., Oei, L., Albagha, O. M., Amin, N., Kemp, J. P., Koller, D. L., Li, G., Liu, C. T., Minster, R. L., Moayyeri, A., Vandenput, L., Willner, D., Xiao, S. M., Yerges-Armstrong, L. M., Zheng, H. F., Alonso, N., Eriksson, J., Kammerer, C. M., Kaptoge, S. K., Leo, P. J., Thorleifsson, G., Wilson, S. G., Wilson, J. F., Aalto, V., Alen, M., Aragaki, A. K., Aspelund, T., Center, J. R., Dailiana, Z., Duggan, D. J., Garcia, M., Garcia-Giralt, N., Giroux, S., Hallmans, G., Hocking, L. J., Husted, L. B., Jameson, K. A., Khusainovam, R., Kim, G. S., Kooperberg, C., Koromila, T., Kruk, M., Laaksonen, M., Lacroix, A. Z., Lee, S. H., Leung, P. C., Lewis, J. R., Masi, L., Mencej-Bedrac, S., Nguyen, T. V., Nogues, X., Patel, M. S., Prezelj, J., Rose, L. M., Scollen, S., Siggeirsdottir, K., Smith, A. V., Svensson, O., Trompet, S., Trummer, O., van Schoor, N. M., Woo, J., Zhu, K., Balcells, S., Brandi, M. L., Buckley, B. M., Cheng, S., Christiansen, C., Cooper, C., Dedoussis, G., Ford, I., Frost, M., Goltzman, D., Gonzalez-Macias, J., Kahonen, M., Karlsson, M., Khusnutdinova, E., Koh, J. M., Kollia, P., Langdahl, B. L., Leslie, W. D., Lips, P., Ljunggren, O., Lorenc, R. S., Marc, J., Mellstrom, D., Obermayer-Pietsch, B., Olmos, J. M., Pettersson-Kymmer, U., Reid, D. M., Riancho, J. A., Ridker, P. M., Rousseau, F., Slagboom, P. E., Tang, N. L., Urreizti, R., Van Hul, W., Viikari, J., Zarrabeitia, M. T., Aulchenko, Y. S., Castano-Betancourt, M., Grundberg, E., Herrera, L., Ingvarsson, T., Johannsdottir, H., Kwan, T., Li, R., Luben, R., Medina-Gomez, C., Palsson, S. T., Reppe, S., Rotter, J. I., Sigurdsson, G., van Meurs, J. B., Verlaan, D., Williams, F. M., Wood, A. R., Zhou, Y., Gautvik, K. M., Pastinen, T., Raychaudhuri, S., Cauley, J. A., Chasman, D. I., Clarkm, G. R., Cummings, S. R., Danoy, P., Dennison, E. M., Eastell, R., Eisman, J. A., Gudnason, V., Hofman, A., Jackson, R. D., Jones, G., Jukema, J. W., Khaw, K. T., Lehtimaki, T., Liu, Y., Lorentzon, M., McCloskey, E., Mitchell, B. D., Nandakumar, K., Nicholson, G. C., Oostra, B. A., Peacock, M., Pols, H. A., Prince, R. L., Raitakari, O., Reid, I. R., Robbins, J., Sambrook, P. N., Sham, P. C., Shuldiner, A. R., Tylavsky, F. A., van Duijn, C. M., Wareham, N. J., Cupples, L. A., Econs, M. J., Evans, D. M., Harris, T. B., Kung, A. W., Psaty, B. M., Reeve, J., Spector, T. D., Streeten, E. A., Zillikens, M. C., Thorsteinsdottir, U., Ohlsson, C., Karasik, D., Richards, J. B., Brown, M. A., Stefansson, K., Uitterlinden, A. G., Ralston, S. H., Ioannidis, J. P., Kiel, D. P. & Rivadeneira, F. (2012). Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genetics* **44**, 491–501.

Flanagan, J. M., Sheehan, V., Linder, H., Howard, T. A., Wang, Y. D., Hoppe, C. C., Aygun, B., Adams, R. J., Neale, G. A. & Ware, R. E. (2013). Genetic mapping and exome sequencing identify 2 mutations associated with stroke protection in pediatric patients with sickle cell anemia. *Blood* **121**, 3237–3245.

Friedl, K. E., Nuovo, J. A., Patience, T. H. & Dettori, J. R. (1992). Factors associated with stress fracture in young army women: indications for further research. *Military Medicine* **157**, 334–338.

Frusztajer, N. T., Dhuper, S., Warren, M. P., Brooks-Gunn, J. & Fox, R. P. (1990). Nutrition and the incidence of stress fractures in ballet dancers. *American Journal of Clinical Nutrition* **51**, 779–783.

Garcia-Ibarbia, C., Perez-Nunez, M. I., Olmos, J. M., Valero, C., Perez-Aguilar, M. D., Hernandez, J. L., Zarrabeitia, M. T., González-Macías, J. & Riancho, J. A. (2013). Missense polymorphisms of the WNT16 gene are associated with bone mass, hip geometry and fractures. *Osteoporosis International* **24**, 2449–2454.

Giladi, M., Milgrom, C., Kashtan, H., Stein, M., Chisin, R. & Dizian, R. (1986). Recurrent stress fractures in military recruits. One-year follow-up of 66 recruits. *Journal of Bone and Joint Surgery. British Volume* **68**, 439–441.

Giladi, M., Milgrom, C., Simkin, A., Stein, M., Kashtan, H., Margulies, J., Rand, N., Chisin, R., Steinberg, R., Aharonson, Z., Kedem, R. & Frankel, V. H. (1987). Stress fractures and tibial bone width. A risk factor. *Journal of Bone and Joint Surgery. British Volume* **69**, 326–329.

Giladi, M., Milgrom, C., Simkin, A. & Danon, Y. (1991). Stress fractures. Identifiable risk factors. *American Journal of Sports Medicine* **19**, 647–652.

Givon, U., Friedman, E., Reiner, A., Vered, I., Finestone, A. & Shemer, J. (2000). Stress fractures in

the Israeli Defense Forces from 1995 to 1996. *Clinical Orthopaedics and Related Research* **373**, 227–232.

Hallel, T., Amit, S. & Segal, D. (1976). Fatigue fractures of tibial and femoral shaft in soldiers. *Clinical Orthopaedics and Related Research* **118**, 35–43.

Hod, N., Ashkenazi, I., Levi, Y., Fire, G., Drori, M., Cohen, I., Bernstine, H. & Horne, T. (2006). Characteristics of skeletal stress fractures in female military recruits of the Israel defense forces on bone scintigraphy. *Clinical Nuclear Medicine* **31**, 742–749.

Jones, B. H., Bovee, M. W., Harris, J. M. 3rd & Cowan, D. N. (1993). Intrinsic risk factors for exercise-related injuries among male and female army trainees. *American Journal of Sports Medicine* **21**, 705–710.

Jurinke, C., van den Boom, D., Cantor, C. R. & Koster, H. (2002). The use of MassARRAY technology for high throughput genotyping. *Advance in Biochemical Engineering/Biotechnology* **77**, 57–74.

Karasik, D., Shimabuku, N. A., Zhou, Y., Zhang, Y., Cupples, L. A., Kiel, D. P. & Demissie, S. (2008). A genome wide linkage scan of metacarpal size and geometry in the Framingham Study. *American Journal of Human Biology* **20**, 663–670.

Karasik, D., Cheung, C. L., Zhou, Y., Cupples, L. A., Kiel, D. P. & Demissie, S. (2012). Genome-wide association of an integrated osteoporosis-related phenotype: is there evidence for pleiotropic genes? *Journal of Bone Mineral Research* **27**, 319–330.

Knapik, J., Montain, S. J., McGraw, S., Grier, T., Ely, M. & Jones, B. H. (2012). Stress fracture risk factors in basic combat training. *International Journal of Sports Medicine* **33**, 940–946.

Korvala, J., Hartikka, H., Pihlajamaki, H., Solovieva, S., Ruohola, J. P., Sahi, T., Barral, S., Ott, J., Ala-Kokko, L. & Mannikko, M. (2010). Genetic predisposition for femoral neck stress fractures in military conscripts. *BMC Genetics* **11**, 95.

Lee, S. H., Baek, J. R., Han, S. B. & Park, S. W. (2005). Stress fractures of the femoral diaphysis in children: a report of 5 cases and review of literature. *Journal of Pediatric Orthopaedics* **25**, 734–738.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Mattila, V. M., Niva, M., Kiuru, M. & Pihlajamaki, H. (2007). Risk factors for bone stress injuries: a follow-up study of 102,515 person-years. *Medicine and Science in Sports and Exercise* **39**, 1061–1066.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303.

Mirina, A., Atzmon, G., Ye, K. & Bergman, A. (2012). Gene size matters. *PloS ONE* **7**, e49093.

Moran, D. S., Evans, R. K. & Hadad, E. (2008). Imaging of lower extremity stress fracture injuries. *Sports Medicine* **38**, 345–356.

Murray, S. R., Reeder, M. T., Udermann, B. E. & Pettitt, R. W. (2006). High-risk stress fractures: pathogenesis, evaluation, and treatment. *Comprehensive Therapy* **32**, 20–25.

Ng, P. C. & Henikoff, S. (2002). Accounting for human polymorphisms predicted to affect protein function. *Genome Research* **12**, 436–446.

Paternoster, L., Ohlsson, C., Sayers, A., Vandenput, L., Lorentzon, M., Evans, D. M. & Tobias, J. H. (2010). OPG and RANK polymorphisms are both associated with cortical bone mineral density: findings from a meta-analysis of the Avon longitudinal study of parents and children and Gothenburg osteoporosis and obesity determinants cohorts. *Journal of Clinical Endocrinology and Metabolism* **95**, 3940–3948.

Paternoster, L., Lorentzon, M., Lehtimaki, T., Eriksson, J., Kahonen, M., Raitakari, O., Laaksonen, M., Sievanen, H., Viikari, J., Lyytikainen, L. P., Mellstrom, D., Karlsson, M., Ljunggren, O., Grundberg, E., Kemp, J. P., Sayers, A., Nethander, M., Evans, D. M., Vandenput, L., Tobias, J. H. & Ohlsson, C. (2013). Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure. *PLoS Genetics* **9**, e1003247.

Ralston, S. H. & Uitterlinden, A. G. (2010). Genetics of osteoporosis. *Endocrine Reviews* **31**, 629–662.

Schwellnus, M. P., Jordaan, G. & Noakes, T. D. (1990). Prevention of common overuse injuries by the use of shock absorbing insoles. A prospective study. *American Journal of Sports Medicine* **18**, 636–641.

Singer, A., Ben-Yehuda, O., Ben-Ezra, Z. & Zaltzman, S. (1990). Multiple identical stress fractures in monozygotic twins. Case report. *Journal of Bone and Joint Surgery. American Volume* **72**, 444–445.

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W. 3rd, Kondrashov, A. S. & Bork, P. (2001). Prediction of deleterious human alleles. *Human Molecular Genetics* **10**, 591–597.

Takahata, M., Iwasaki, N., Nakagawa, H., Abe, Y., Watanabe, T., Ito, M., Majima, T. & Minami, A. (2007). Sialylation of cell surface glycoconjugates is essential for osteoclastogenesis. *Bone* **41**, 77–86.

Valimaki, V. V., Alfthan, H., Lehmuskallio, E., Loyttyniemi, E., Sahi, T., Suominen, H. & Valimaki, M. J. (2005). Risk factors for clinical stress fractures in male military recruits: a prospective cohort study. *Bone* **37**, 267–273.

Villar, V. A., Jones, J. E., Armando, I., Palmes-Saloma, C., Yu, P., Pascua, A. M., Keever, L., Arnaldo, F. B., Wang, Z., Luo, Y., Felder, R. A. & Jose, P. A. (2009). G protein-coupled receptor kinase 4 (GRK4) regulates the phosphorylation and function of the dopamine D3 receptor. *Journal of Biological Chemistry* **284**, 21425–21434.

Yan, J. & Aldrich, R. W. (2012). BK potassium channel modulation by leucine-rich repeat-containing proteins. *Proceedings of the National Academy of Science of the United States of America* **109**, 7917–7922.

Yanovich, R., Milgrom, R., Friedman, E. & Moran, D. S. (2011). Androgen receptor CAG repeat size is associated with stress fracture risk: a pilot study. *Clinical Orthopaedics and Related Research* **469**, 2925–2931.

Yanovich, R., Friedman, E., Milgrom, R., Oberman, B., Freedman, L. S. & Moran, D. S. (2012). Candidate gene analysis in Israeli soldiers with stress fractures. *Journal of Sport Science and Medicine* **11**, 147–155.

Zhang, F., Qiu, T., Wu, X., Wan, C., Shi, W., Wang, Y., Chen, J. G., Wan, M., Clemens, T. L. & Cao, X. (2009). Sustained BMP signaling in osteoblasts stimulates bone formation by promoting angiogenesis and osteoblast

differentiation. *Journal of Bone and Mineral Research* **24**, 1224–1233.

Zhao, L. J., Liu, X. G., Liu, Y. Z., Liu, Y. J., Papasian, C. J., Sha, B. Y., Pan, F., Guo, Y. F., Wang, L., Yan, H., Xiong, D. H., Tang, Z. H., Yang, T. L., Chen, X. D., Guo, Y., Li, J., Shen, H., Zhang, F., Lei, S. F., Recker, R. R. & Deng, H. W. (2010). Genome-wide association study for femoral neck bone geometry. *Journal of Bone and Mineral Research* **25**, 320–329.

Zheng, H. F., Tobias, J. H., Duncan, E., Evans, D. M., Eriksson, J., Paternoster, L., Yerges-Armstrong, L. M., Lehtimaki, T., Bergstrom, U., Kahonen, M., Leo, P. J., Raitakari, O., Laaksonen, M., Nicholson, G. C., Viikari, J., Ladouceur, M., Lyytikainen, L. P., Medina-Gomez, C., Rivadeneira, F., Prince, R. L., Sievanen, H., Leslie, W. D., Mellstrom, D., Eisman, J. A., Moverare-Skrtic, S., Goltzman, D., Hanley, D. A., Jones, G., St Pourcain, B., Xiao, Y., Timpson, N. J., Smith, G. D., Reid, I. R., Ring, S. M., Sambrook, P. N., Karlsson, M., Dennison, E. M., Kemp, J. P., Danoy, P., Sayers, A., Wilson, S. G., Nethander, M., McCloskey, E., Vandenput, L., Eastell, R., Liu, J., Spector, T., Mitchell, B. D., Streeten, E. A., Brommage, R., Pettersson-Kymmer, U., Brown, M. A., Ohlsson, C., Richards, J. B. & Lorentzon, M. (2012). WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. *PLoS Genetics* **8**, e1002745.

Zwas, S. T., Elkanovitch, R. & Frank, G. (1987). Interpretation and classification of bone scintigraphic findings in stress fractures. *Journal of Nuclear Medicine* **28**, 452–457.