SCIENTIFIC REPORTS

Corrected: Author Correction

OPEN

Whole exome sequencing for the identification of CYP3A7 variants associated with tacrolimus concentrations in kidney transplant patients

Minji Sohn[1], Myeong Gyu Kim[1,2], Nayoung Han[1], In-Wha Kim[1], Jungsoo Gim[3], Sang-Il Min[4], Eun Young Song[5], Yon Su Kim[6], Hun Soon Jung[7], Young Kee Shin[1,8], Jongwon Ha[4] & Jung Mi Oh [ID][1]

The purpose of this study was to identify genotypes associated with dose-adjusted tacrolimus trough concentrations (C_0/D) in kidney transplant recipients using whole-exome sequencing (WES). This study included 147 patients administered tacrolimus, including seventy-five patients in the discovery set and seventy-two patients in the replication set. The patient genomes in the discovery set were sequenced using WES. Also, known tacrolimus pharmacokinetics-related intron variants were genotyped. Tacrolimus C_0/D was log-transformed. Sixteen variants were identified including novel *CYP3A7* rs12360 and rs10211 by ANOVA. *CYP3A7* rs2257401 was found to be the most significant variant among the periods by ANOVA. Seven variants including *CYP3A7* rs2257401, rs12360, and rs10211 were analyzed by SNaPshot in the replication set and the effects on tacrolimus C_0/D were verified. A linear mixed model (LMM) was further performed to account for the effects of the variants and clinical factors. The combined set LMM showed that only *CYP3A7* rs2257401 was associated with tacrolimus C_0/D after adjusting for patient age, albumin, and creatinine. The *CYP3A7* rs2257401 genotype variant showed a significant difference on the tacrolimus C_0/D in those expressing *CYP3A5*, showing its own effect. The results suggest that *CYP3A7* rs2257401 may serve as a significant genetic marker for tacrolimus pharmacokinetics in kidney transplantation.

Tacrolimus is an immunosuppressant drug widely used in most organ transplants[1]. Tacrolimus has successfully decreased the rejection rate and improved the outcome of many transplants[2]. However, individualization of tacrolimus therapy remains a challenge owing to the wide range of inter-individual variations in its pharmacokinetics (PK) and its narrow therapeutic index[3]. Accordingly, therapeutic drug monitoring (TDM) of tacrolimus leads to better clinical outcomes to prevent adverse effects and decrease the risk of allograft rejection in clinical settings.

Various factors have been reported to affect tacrolimus PK, such as concomitant drugs, genotypes, diet, and clinical values[4,5]. Notably, the *CYP3A5* rs776746 single-nucleotide polymorphism (SNP) found in intron 3 (6986A>G, *CYP3A5*3* allele) is known to play a major role in the expression of the CYP3A5 enzyme via abnormal mRNA splicin[6], which influences tacrolimus absorption and metabolism. Several guidelines have been

[1]College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, Republic of Korea. [2]Graduate school of clinical pharmacy, CHA University, Gyeonggi-do, Republic of Korea. [3]Department of Biomedical Science, Chosun University, Gwangju, Republic of Korea. [4]Department of Surgery, Seoul National University Hospital, Seoul, Republic of Korea. [5]Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Republic of Korea. [6]Kidney Research Institute and Department of Medical Science, Seoul National University College of Medicine, Seoul, Republic of Korea. [7]R&D center, ABION, Inc., Seoul, Republic of Korea. [8]Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, 08826, Republic of Korea. Jongwon Ha and Jung Mi Oh contributed equally. Correspondence and requests for materials should be addressed to J.H. (email: jwhamd@snu.ac.kr) or J.M.O. (email: jmoh@snu.ac.kr)

	Discovery patients (N = 75)	Replication patients (N = 72)	P value
Age (years)	46.2 ± 13.0	48.0 ± 12.4	0.386
Male, N (%)	46 (61.3)	48 (66.7)	0.759
Body weight (kg)	60.8 ± 11.8	60.3 ± 9.3	0.778
Deceased donor, N (%)	44 (58.7)	41 (53.9)	0.354
Origin of kidney disease, N (%)			
Diabetes	19 (25.3)	13 (18.1)	0.112
Hypertension	6 (8.0)	10 (13.9)	0.147
Glomerulonephritis	26 (34.7)	26 (36.1)	0.603
PKD	7 (9.3)	7 (9.7)	0.512
Others	7 (9.3)	4 (5.6)	0.544
Unknown/CGN	10 (13.3)	12 (16.7)	0.470
Clinical variables			
Hematocrit (%)	34.8 ± 3.9	33.9 ± 3.6	0.114
Total cholesterol (mg/dL)	151.8 ± 36.1	152.3 ± 40.2	0.928
Albumin (g/dL)	3.5 ± 0.5	3.5 ± 0.6	0.995
Total bilirubin (mg/dL)	0.7 ± 0.3	0.7 ± 0.4	0.913
Alanine transaminase (U/L)	15.2 ± 10.2	16.3 ± 10.3	0.494

Table 1. Baseline demographic characteristics for included subjects. PKD, polycystic kidney disease; CGN, chronic glomerulonephritis; Data are presented as a number with percentage for categorical variables and a mean with standard deviation for continuous variables.

implemented to compensate for this; for example, the Clinical Pharmacogenetics Implementation Consortium guidelines recommend a 1.5–2-fold higher starting dose for CYP3A5 expressers, such as *CYP3A5*1* carriers, compared to CYP3A5 non-expressers, such as *CYP3A5*3* carriers[7]. Nevertheless, efforts are being made to discover the genetic determinants of tacrolimus PK because *CYP3A5*3* is believed to account for 40–50% of variations in tacrolimus dosage requirements[8], and the correct dosage is still undetermined for a large proportion of patients. Recently, several other variants in *CYP3A4*[9], *ABCB1*[10], *POR*[11], *NR1I2*[12], and *SUMO4*[13] genes have been found to affect tacrolimus PK. However, these genetic variants do not explain the substantial variability observed for tacrolimus PK.

Genome-wide screening studies constitute a potential new approach to identify novel SNPs. Next–generation sequencing (NGS) has proven highly successful in identifying novel pathological genotypes[14,15]. Hence, this new technology is expected to reveal novel genetic variants to successfully predict tacrolimus PK. Additionally, a patient's clinical condition can affect tacrolimus PK[2]; hence, clinical factors or laboratory variables associated with tacrolimus PK are required to predict the tacrolimus trough whole-blood concentration (C_0). Therefore, this study aimed to identify genotypes via NGS and clinical factors associated with tacrolimus C_0 after kidney transplantation (KT).

Results

Patient characteristics in the discovery and replication sets.
The baseline demographic characteristics are presented in Table 1. There were no significant differences in age (46.2 ± 13.0 vs. 48.0 ± 12.4 years old, $P = 0.386$), body weight (60.8 ± 11.8 vs. 60.3 ± 9.3 kg, $P = 0.778$), or sex (61.3% vs. 66.7% male patients, $P = 0.759$) between the discovery and replication sets. Drugs with potential effects on tacrolimus PK, such as CYP3A inhibitors or inducers, were not administered.

Sequencing and alignment quality.
After sequence mapping, the mean coverage depth was 57.78x (41.09–105.5), and 87.87% (77.87–96.09%) of target regions showed >20x coverage (Supplementary Table S1). After variant calling, 293,531 variants were identified and used for association. The data quality and quantity are described in Supplementary Table S1.

Variants associated with tacrolimus trough concentrations using whole-exome and intron data.
In total, 2,900 tacrolimus C_0 measurements were collected from patients during the first year after transplantation in the discovery set. The median C_0 per patient was 37 points, with a range from 28–94 points. After analyzing the association between the variants and the tacrolimus dose-adjusted trough concentration (C_0/D) using ANOVA, genes such as *NR1I2* on chromosome 3 and *PTCD1*, *CPSF4*, *ZNF789*, *ZKSCAN5*, *FAM200A*, *ZSCAN25*, *CYP3A5*, *CYP3A7*, and *CYP3A4* on chromosome 7 were strongly associated with tacrolimus C_0/D (Fig. 1 and Supplementary Figure S1) at one or more time points. All the variants conformed with the Hardy-Weinberg Equilibrium ($P > 0.001$) and matched those reported for East Asian populations (Supplementary Table S2). Sixteen variants were significantly associated with tacrolimus C_0/D after FDR ($P < 0.05$) from 3 days to 1 year after transplantation at each time point (Fig. 2 and Supplementary Table S3). The *CYP3A7* rs2257401 C to G substitution was the most significant variant for increasing the tacrolimus C_0/D in all periods and was the only variant that strongly associated with tacrolimus C_0/D on day 3 after transplantation (between $P = 1.74 \times 10^{-7}$ and $P = 0.0138$, Supplementary Table S3 and Fig. 3). The fourteen other variants associated with tacrolimus C_0/D were
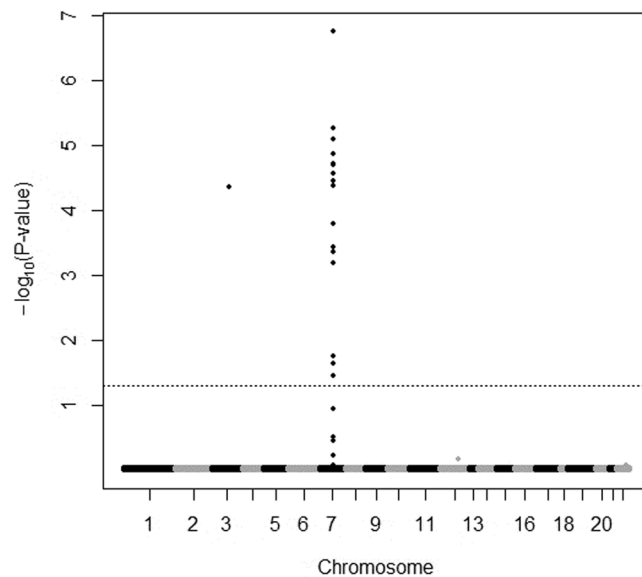
Figure 1. Manhattan plots of variants associated with dose-adjusted tacrolimus trough concentrations on day 7 after transplantation. The dotted horizontal line shows the cutoff of $P = 0.05$ after false discovery rate correction.
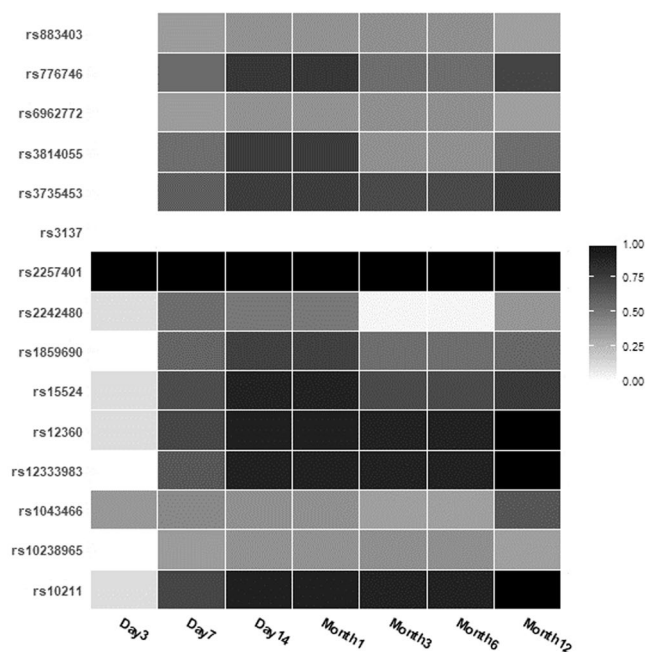

Figure 2. Heat map plots of variants commonly associated with dose-adjusted tacrolimus trough concentrations from day 3 to 1 year after transplantation. The color was scaled based on an adjusted P value of variants associated with tacrolimus trough levels (black, highly significant association; white, no significant association).

found at multiple times between day 7 and 1 year (*CPSF1* rs883403 and rs1043466; *ZNF789* rs6962772; *FAM200A* rs10238965; *ZSCAN25* rs1859690 and rs3735453; *CYP3A5* rs15524 and rs776746; *CYP3A7* rs10211, rs12360, and rs2257401; *CYP3A4* rs12333983 and rs2242480; and *NR1I2* rs3814055). The mean tacrolimus C_0 in relation to these seven variants in drug-metabolizing enzymes, including *CYP3A4*, *CYP3A5*, and *CYP3A7* and the nuclear receptor gene *NR1I2*, at different time points after transplantation are presented in Supplementary Table S4.

Association of haplotypes with tacrolimus trough concentrations.   A haplotype distribution was constructed for thirteen variants in seven genes associated with tacrolimus C_0/D, including *CPSF1*, *ZNF789*, *FAM200A*, *ZSCAN25*, *CYP3A5*, *CYP3A7*, and *CYP3A4* on chromosome 7. Linkage disequilibrium (LD) structures were designated by the r^2 values shown in Fig. 4. One haplotype block consisted of *CPSF1* rs883403 and
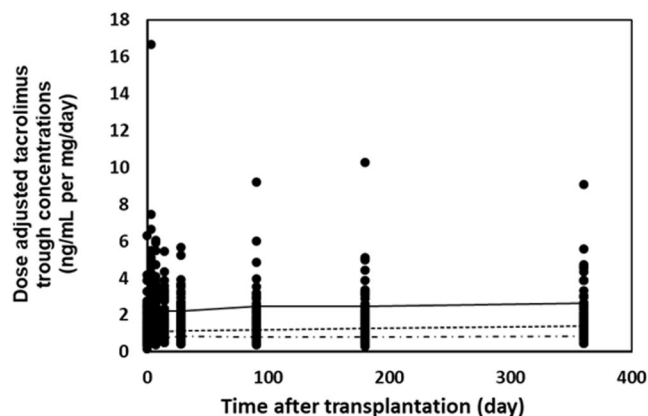
Figure 3. Dose-adjusted tacrolimus trough concentrations in patients after kidney transplantation. Solid, dotted, and dashed lines indicated the mean of dose-adjusted tacrolimus trough concentrations in carriers with *CYP3A7* rs2257401 GG, CG, and CC genotypes, respectively.
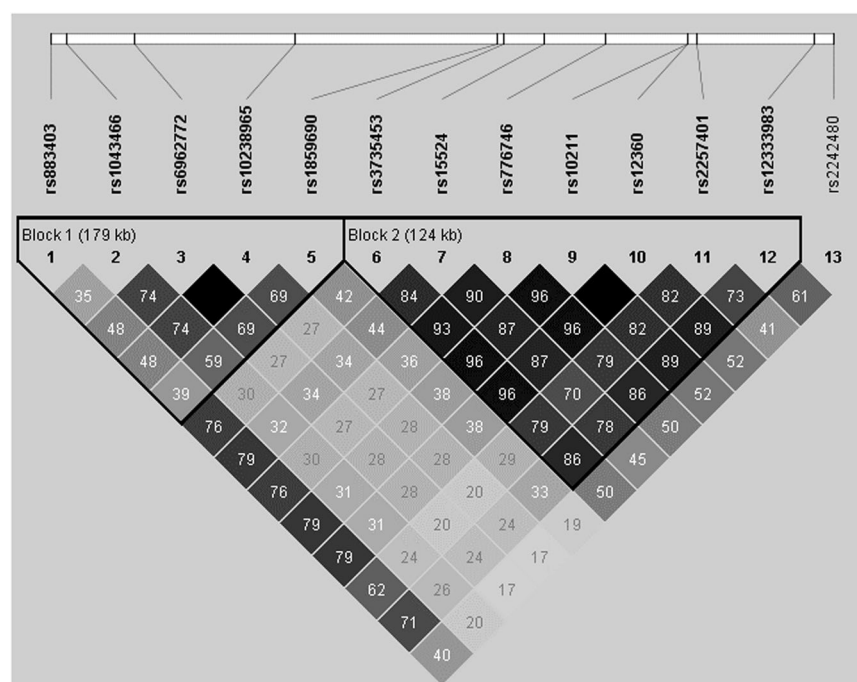

Figure 4. Haplotype plots of variants on chromosome 7 commonly associated with dose-adjusted tacrolimus trough concentrations from day 7 to 1 year after transplantation.

rs1043466, *ZNF789* rs6962772 and *FAM200A* rs10238965, and *ZSCAN25* rs1859690; the other consisted of *ZSCAN25* rs3735453 and *CYP3A5* rs15524 and rs776746; *CYP3A7* rs10211, rs12360, and rs2257401; and *CYP3A4* rs12333983. The *ZNF789* rs6962772, *FAM200A* rs10238965, and *CYP3A7* rs10211 and rs12360 SNPs displayed complete linkage; therefore, *ZNF789* rs6962772 and *CYP3A7* rs10211 were not used in further analyses. There was a high degree of LD between *CYP3A7* rs2257401 and *CYP3A5* rs776746 ($r^2 = 0.79$) and a moderate degree of LD between *CYP3A4* rs2242480 and *CYP3A5* rs776746 ($r^2 = 0.50$). The diplotype frequencies of *CYP3A5* rs15524 and rs776746, *CYP3A7* rs10211 and rs2257401, and *CYP3A4* rs12333983, including CAGCA-CAGCA, TGAGT-TGAGT, and others, are summarized in Supplementary Table S5. Haplotype combinations showed that the CAGCA-CAGCA, TGAGT-TGAGT, and other diplotypes were observed at frequencies of 9.3%, 45.3%, and 44.2%, respectively. The TGAGT-TGAGT diplotype was strongly associated with increased tacrolimus C_0/D compared to CAGCA-CAGCA at different time points during the first year after transplantation (Supplementary Table S6).

Association of genotypes and clinical factors with tacrolimus trough concentrations.

Biochemical data including the hematocrit, serum albumin, total cholesterol, total bilirubin, alanine

Variables	Univariate		Multivariate	
	Estimate	95% CI	Estimate	95% CI
Discovery set (N = 75)				
Age	1.018	1.010 − 1.027	1.014	1.009−1.020
Serum creatinine	1.064	1.025–1.105	1.053	1.015–1.093
Serum albumin	0.942	0.890–0.997	0.959	0.906–1.015
CYP3A5 rs776746 A>G	1.654	1.444–1.894	1.201	0.639–1.080
CYP3A7 rs2257401 C>G	1.736	1.539–1.959	1.784	1.401–2.271
CYP3A4 rs2242480 T>C	1.733	1.436–2.090	1.227	1.022–1.474
Replication set (N = 72)				
Age	1.004	0.994–1.014		
Serum creatinine	1.066	1.031–1.102	1.055	1.020–1.091
Serum albumin	0.925	0.878–0.975	0.946	0.897–0.998
CYP3A5 rs776746 A>G	1.867	1.634–2.133	2.071	1.464–2.931
CYP3A7 rs2257401 C>G	1.740	1.495–2.026	0.886	0.634–1.237
CYP3A4 rs2242480 T>C	1.552	1.299–1.854	1.024	0.858–1.221
Combined set (N = 147)				
Age	1.011	1.004–1.018	1.010	1.005–1.015
Serum creatinine	1.066	1.039–1.093	1.057	1.030–1.084
Serum albumin	0.935	0.899–0.971	0.956	0.919–0.994
CYP3A5 rs776746 A>G	1.750	1.591–1.925	1.245	0.993–1.561
CYP3A7 rs2257401 C>G	1.727	1.568–1.902	1.328	1.074–1.642
CYP3A4 rs2242480 T>C	1.634	1.854–1.441	1.134	0.993–1.294

Table 2. Results from the final linear mixed model for tacrolimus dose-adjusted trough concentrations in kidney transplant recipients.

transaminase, and creatinine were measured during the study period (Supplementary Table S7). A mixed effect model analysis was performed to further identify genetic variants and clinical factors. Patient age at transplantation and serum creatinine were significantly positively associated with tacrolimus C_0/D in the discovery set ($P < 0.001$, Table 2). The C–G substitution rate for *CYP3A7* rs2257401 and T–C substitution rate for *CYP3A4* rs2242480 were significantly increased in patients with a dose-adjusted tacrolimus C_0/D after adjusting for age and serum creatinine levels ($P < 0.001$).

Validation of genotypes and clinical factors associated with tacrolimus trough concentrations in the replication set. In total, 2,940 tacrolimus C_0 measurements were obtained from patients during the first year after transplantation in the replication set. The median C_0 per patient was 34 points (ranging from 27–77 points). Seven variants associated with tacrolimus C_0/D in the discovery set, including *CYP3A5* rs15524 and rs776746; *CYP3A7* rs10211, rs12360, and rs2257401; and *CYP3A4* rs12333983 and rs2242480, were genotyped by SNaPshot assay in the replication set. The allele frequencies of these variants are presented in Supplementary Table S8 and were comparable with the frequencies in the discovery set. Mixed effect model analysis showed that serum albumin and creatinine were positively associated with tacrolimus C_0/D in the replication set ($P < 0.05$, Table 2). After adjusting for serum albumin and creatinine, only the A to G substitution rate of *CYP3A5* rs776746 was significantly increased in patients with dose-adjusted tacrolimus C0/D ($P < 0.001$). The mean tacrolimus C0/D based on the *CYP3A5* rs776746 genotypes at different time points are summarized in Supplementary Table S9.

Genotypes and clinical factors associated with tacrolimus trough concentrations in the combined set. A combined analysis of both the discovery and replication sets showed that only the *CYP3A7* rs2257401 variant was associated with tacrolimus C_0/D after adjusting for patient age, serum albumin, and creatinine ($P < 0.05$, Table 2). Next, the effect of the *CYP3A7* rs2257401 and *CYP3A5* rs776746 variants on the tacrolimus C_0/D in patients expressing *CYP3A5* (*CYP3A5*1/*1* or *CYP3A5*1/*3*) and not expressing *CYP3A5* (*CYP3A5*3/*3*) was examined. There was a significant difference in tacrolimus C_0/D in carriers of the *CYP3A7* rs2257401 genotype for those expressing CYP3A5 ($P < 0.001$), but no significant difference in tacrolimus C_0/D was found in carriers of the *CYP3A7* rs2257401 genotype for those not expressing *CYP3A5* ($P = 0.180$, Table 3). The *CYP3A7* rs2257401 CC genotype was not found in patients not expressing *CYP3A5*.

Discussion

To our knowledge, this is the first study to perform a combined whole-exon and intron association analysis with clinical factors associated with tacrolimus C_0/D in Korean KT recipients. Sixteen variants including *CYP3A* family genes with novel SNPs, *CYP3A7* rs12350 and rs10211, were identified to affect tacrolimus C_0/D. The *CYP3A7* rs2257401 SNP variant was most significantly associated with tacrolimus C_0/D and related after adjusting for patient age, serum albumin, and creatinine levels.

CYP3A5	rs2257401	N	Daily dose adjusted tacrolimus trough concentrations (ng/mL per mg/day)						
			Day 3	Day 7	Day 14	Month 1	Month 3	Month 6	Month 12
Expressers*	GG or CG	52	1.61 ± 0.99	1.23 ± 0.75	1.05 ± 0.43	1.15 ± 0.69	1.20 ± 0.78	1.17 ± 0.65	1.33 ± 0.66
	CC	15	0.99 ± 0.46	0.70 ± 0.35	0.74 ± 0.26	0.82 ± 0.35	0.78 ± 0.28	0.74 ± 0.25	0.82 ± 0.25
Nonexpressers	GG	71	3.26 ± 2.29	2.58 ± 1.39	2.17 ± 1.09	2.09 ± 0.97	2.25 ± 1.23	2.18 ± 1.33	2.37 ± 1.36
	CG	9	2.11 ± 1.61	1.80 ± 1.23	1.83 ± 1.05	2.04 ± 1.29	1.84 ± 0.92	2.00 ± 0.87	2.16 ± 1.29

Table 3. Tacrolimus dose-adjusted trough concentrations in *CYP3A7* rs2257401 genotypes classified by different patients expressing *CYP3A5*. Data are presented as a mean with standard deviation. ***P < 0.05 analyzed by a mixed effect model after adjusted by patient age, serum albumin, and creatinine. *CYP3A5* expressers are carriers with *CYP3A5*1/*1 or CYP3A5*1/*3*.

CYP3A7 is a major CYP enzyme in fetal livers. However, it is rapidly downregulated within the first few years postpartum[16]. Sim *et al.* reported that CYP3A7 protein expression is higher than CYP3A5 expression in 10% of adult livers, which may be relevant for the metabolism of various substrates[17]. The underlying mechanism is assumed to involve a C–G substitution in *CYP3A7* rs2257401, which causes a Thr/Arg amino acid substitution, thereby decreasing enzyme activity in an *in vitro* study using human HEK293 cells[18]. However, few studies have questioned the role of the CYP3A7 enzyme. Recombinant CYP3A7 supersomes showed lower metabolic activity than CYP3A4 and CYP3A5 supersomes[19]. The metabolic velocity of CYP3A7 for tacrolimus was lower than that of CYP3A5, whereas the K_m value of CYP3A7 was higher than that of CYP3A5 in a baculovirus-expressed CYP system[20].

For calcineurin inhibitors, a few studies have described the effect of *CYP3A7*1C*, a promoter variant affecting their metabolism; however, their results were inconclusive[21–23]. Carriers of the *CYP3A7*1C* allele maintained a higher level of expression into adulthood and required higher doses of cyclosporine[21,22]. Nonetheless, Elens *et al.* reported that this polymorphism had no effect on tacrolimus C_0/D and dose requirements in liver transplant recipients[23]. *CYP3A7* rs2257401 has been reportedly associated with the area under the concentration-time curve of tacrolimus in healthy Korean subjects[12]. *CYP3A7* rs12360 and rs10211 are SNPs in distance of 100 bp in the 3′-UTR region. Few microRNAs are reported to bind with rs12360 suggesting to modify gene expression, whereas none are reported to bind rs10211[24].

The *CYP3A7* rs2257401 allele frequency was demonstrated to differ substantially between ethnic groups. The minor allele frequency (MAF) of the rs2257401 C allele was different in Tanzanian (MAF = 0.62), Chinese (MAF = 0.28), Saudi Arabian (MAF = 0.17), and Caucasian (MAF = 0.08) populations[18]. The MAF of the *CYP3A7* rs2257401 genotype in the present study was 30.7%, which was similar to findings from a previous study on healthy Korean subjects[21–23,25]. *CYP3A7* rs2257401 variation was reported to be higher in individuals of Asian descent; this information showed that this genotype might be a useful biomarker, as it was the major significant variant associated with tacrolimus C_0/D in the Korean population.

The present results showed that patient age was associated with an increase in the tacrolimus C_0/D, which was consistent with previous findings[8,26]. Additionally, serum albumin levels of patients were positively associated with an increase in the tacrolimus C_0/D. A previous study reported the influence of albumin levels on tacrolimus clearance[27]. Serum albumin levels in the present study were increased soon after transplantation; however, they returned to baseline within 1 month. Restoration of albumin levels could increase the levels of bound tacrolimus in the blood and thus reduce tacrolimus clearance. Furthermore, an apparent correlation between serum creatinine levels and tacrolimus C_0/D was previously observed[28], which was consistent with the present findings. Thus, any increase in systemic exposure to tacrolimus would increase serum creatinine levels[29].

In the linear mixed model (LMM), the variants selected from each genotype and clinical values were used as a covariate. All were associated with tacrolimus C_0/D for each; however, only one variant was retained in the final model. The final LMM included *CYP3A7* rs2257401 to best explain tacrolimus C_0/D in the discovery set ($P < 0.05$ using a *chi-square* test between the model with rs2257401 and the model with rs776746). *CYP3A5* rs776746 was selected in the final LMM in the replication set (non-significant *chi-square* test between two models; which indicates that model fitness's are similar). Such results suggested that even though the final LMM in discovery set was not confirmed, the role of *CYP3A7* rs2257401 was still important. The difference of the results may have occurred because *CYP3A5* rs776746 variants were detected more frequently in the replication set. This explanation coincides with previous reports that the *CYP3A7* is strongly but not completely linked to *CYP3A5* rs776746[12,30], which means a larger study is required to verify the importance of *CYP3A7* rs2257401 for personalized tacrolimus therapy.

A high degree of LD between *CYP3A7* rs2257401 and *CYP3A5* rs776746 ($r^2 = 0.79$) was observed and reported in Birdwell *et al.*, an another tacrolimus PK study without the effect of *CYP3A7* rs2257401[31]. To determine the sole effect of CYP3A7 in the present study, the recipients were categorized as those characterized by *CYP3A5* expression and those lacking *CYP3A5* expression, and the effect of *CYP3A7* variants on tacrolimus C_0/D was examined in each group. A substantial difference was observed in the C_0/D between patients with and without the *CYP3A7* rs2257401 genotype among those characterized by *CYP3A5* expression, which suggests its own effect on the CYP3A7 enzyme.

In our study of the discovery set, the *CYP3A4*1G* (rs2242480, 20239T>C) variant was found to have a positive association with tacrolimus C_0/D, which was consistent with previous reports[12]. The *CYP3A4* rs2242480 SNP located on intron 10 has been reported to be associated with tacrolimus C_0/D in Asian KT recipients, but the functional role of this polymorphism remains unclear[32,33]. *CYP3A4*1B* (rs2740574, A>G) has been reported

to be associated with tacrolimus C_0/D[34]; however, it was not tested in the current study because no variant was reported in individuals of Asian descent by dbSNP[35].

Although numerous studies have attempted to investigate whether three *ABCB1* polymorphisms, 1236C>T (rs1128503), 2677G>A/T (rs2032582), and 3435C>T (rs1045642), affect tacrolimus PK, few have successfully shown any association[10,27]. In the present study, no associations were observed between tacrolimus C_0/D and these *ABCB1* polymorphisms, which was identical with Hesselink *et al.*[34]. Although the absorption of tacrolimus is affected by P-glycoprotein, the roles of these *ABCB1* polymorphisms remain unclear.

Except *CYP3A* family, other gene variants were also identified. When Bonferroni correction was applied, all the variants except *NR1I3* rs3814055 were also significant (threshold of *P*-value $< 2.5 \times 10^{-6}$, data not shown)[36]. *PTCD1* encodes a mitochondrial protein and rs28495024 was related to distinct mtDNA gene expression in Cohen *et al.*[37]. *CPSF4* rs883403 and rs14043466 have been identified in genome studies on various diseases. However, too many unrelated diseases have been reported, and the exact role of the variant is in question[38]. *ZKSCAN5* is suggested to be involved in transcriptional regulation and rs3137 is reported to bind with microR-NAs, resulting in a modified function[39]. The role of *ZNF789* gene variant has not been reported in any database. For *ZSCAN25*, which is also known as *ZNF498*, a study suggests that its gene region overlaps with *CYP3A5*, but no evidence was presented[40]. Owing to no known functions on the pharmacokinetics of the genes, the genes were not selected for verification in the replication set.

Even though sequencing methods have advanced, NGS is still challenging. Numerous studies have compared sequencers and bioinformatics tools, but the best technique has not yet been concluded[41,42]. In our data, while the average coverage was larger than 55, the fact that 13% of the exome is covered with less than 20x suggests that it would be challenging to call these regions. The percent bases in the target reads was low, but it was similar with previous studies; nonetheless, improvements are still required[43,44].

To our knowledge, the statistical tools to test the power of NGS even with continuous variables, such as drug concentrations rather than disease risk, are not established[45]. When roughly estimated by the Genome Power Calculator[46], the power to detect *CYP3A7* rs2257401 was supposedly over 95%, though it was still low for detecting rare variants. The current research was designed to identify clinically practical genotypes and we expected the results to be meaningful. As the sample size of this study was considered not large enough to identify mutations, those were excluded for validation.

Although the present study clearly identified the relationship between genotypes and tacrolimus trough levels, certain limitations exist. The exact function of *CYP3A7* was not determined; therefore, further studies are required to determine the physiological role of the *CYP3A7* enzyme in drug metabolism. Furthermore, future prospective studies are warranted to adequately characterize the effects of the *CYP3A7* variant and clinical factors on the tacrolimus PK.

In conclusion, the present study reported that *CYP3A7* polymorphisms are strongly associated with individual differences in the tacrolimus C_0/D. Therefore, genotyping of the *CYP3A7* rs2257401 polymorphism may help optimize personalized tacrolimus dosages for KT recipients.

Materials and Methods

Study design and population. The current retrospective observational cohort study was performed in a single center at Seoul National University Hospital (SNUH). Patients aged 18 years or older who underwent KT from January 2007 to September 2014 and were treated with tacrolimus (Prograf, Astellas Pharma Korea, Inc., Seoul, Korea) during the first year after transplantation were included in this study. Patients were excluded if they underwent other organ transplants or if they received desensitization therapy due to the presence of donor-specific antibodies or ABO blood-type incompatibility. A total of 147 patients were included in this study; seventy-five were placed in the discovery set, and seventy-two were placed in the replication set. This study was approved by the ethics committee of SNUH (IRB No. C-1504-009-662) and was performed in accordance with the Guidelines for Good Clinical Practices and the Declaration of Helsinki[47]. Two written informed consent documents, for study participation and genetic testing of blood samples, were obtained from all study subjects. All subjects studied were of Korean ethnicity and no prisoners were included.

Immunosuppressive regimen. Induction therapy included 20 mg preoperative intravenous basiliximab (Simulect, Novartis Pharmaceuticals, East Hanover, NJ, USA) and 4 days postoperative basiliximab or rabbit antithymocyte globulin (Thymoglobulin, Sanofi, Paris, France) for 4–7 days, as well as 0.5 g preoperative intravenous methylprednisolone (Methysol, Alvogen Korea, Seoul, Korea). Subsequent maintenance triple immunosuppressive therapy was employed, including tacrolimus, mycophenolate (Cellcept, Roche, Nutley, NJ, USA; or Myfortic, Novartis, East Hanover, NJ, USA), and steroids. Tacrolimus was prescribed at an initial oral dose of 0.075 mg/kg twice daily from the day before transplantation and was adjusted thereafter according to TDM methods. The daily dose was adjusted to maintain the tacrolimus C_0 at 10–12 ng/mL during the first month after KT, 8–10 ng/mL until 3 months, 6–8 ng/mL until 6 months, and 4–6 ng/mL thereafter. Prednisolone was gradually tapered to 5–10 mg/day before the patients were discharged and 10 mg/day by 2 weeks after discharge. Additionally, 500 mg mycophenolate mofetil or 360 mg mycophenolate sodium was administered every 12 h on the day of transplantation; the dose was adjusted according to the side effects.

Tacrolimus trough concentrations and clinical variable data collection. Ethylenediaminetetraacetic acid-anticoagulated whole blood samples were collected from kidney recipients visiting the department of Surgery just before the morning dose was administered. The tacrolimus C_0 was measured at regular intervals for the first year after transplantation. The tacrolimus C_0 was analyzed in whole blood by liquid-chromatography tandem mass spectrometry (LC-MS/MS) using a Waters 2795 Alliance HT system (Waters Ltd., Watford, UK) and a Quattro micro API tandem mass spectrometer (Micromass, Manchester, UK) as described previously[48]. The

mean tacrolimus C_0 values were calculated for the following periods after KT: days 1–3, days 4–7, days 8–14, and days 15–28, months 1–3, months 4–6, and months 7–12. The clinical information for each patient, including the age at transplantation, sex, body weight, donor source, concomitant medications, and previous history of transplantation, were obtained; the biochemical laboratory variables, including the hematocrit, total serum cholesterol, total bilirubin, albumin, and creatinine, were collected at the time of tacrolimus C_0 sampling at regular intervals for the first year after transplantation.

Genomic DNA isolation and whole exome sequencing.   Genomic DNA (gDNA) was isolated from recipient whole blood using the QuickGene DNA whole blood kit (Kurabo Industries, Osaka, Japan) according to the manufacturer's instructions. The purity and concentration of gDNA was measured using a NanoDrop (Thermo Fisher Scientific, Grand Island, NY, USA). The exons and untranslated regions (UTRs) in the discovery set were genotyped by whole exome sequencing (WES). WES was conducted using 1 μg fragmented gDNA, which was captured using the Sure Select Human All Exon kit V5 + UTRs (Agilent Technologies, Santa Clara, CA, USA) and amplified. The libraries were evaluated for quality and quantified with a High Sensitivity DNA kit using a Bioanalyzer (Agilent Technologies). Enrichment was conducted using an Ion OneTouch ES (Thermo Fisher Scientific). Samples were loaded on two Ion PI chip Kits and sequenced on the Ion Proton System using an Ion PI Hi-Q Sequencing 200 Kit (200-bp read length, Thermo Fisher Scientific). Reads were mapped against the human reference genome (hg19) using the Torrent Mapping Alignment Program version 4.0.6 (Thermo Fisher Scientific). Variant calling was performed by running the Torrent Variant Caller plugin version 4.4.3.3 with the recommended optimized parameters for exome sequencing. Individual SNPs were excluded if they were monomorphic or had a low minor allele frequency (MAF; <1%). Variants were annotated and classified as deleterious by SNVrap (http://jjwanglab.org/snvrap)[49].

Sanger sequencing.   Sanger sequencing confirmed all of the samples with the *CYP3A7* rs2257401 candidate variant (Supplementary Figure S2). Variant-rich regions were amplified by PCR (primer sets are listed in Supplementary Table S10). Purified PCR amplicons were directly sequenced with the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Subsequent analysis was performed using a 3130xl Genetic Analyzer (Thermo Fisher Scientific).

SNaPshot or SNPtype assay.   Non-targeted regions including intronic variants identified by the SureSelect kit in the discovery set were genotyped using the SNaPshot Multiple Kit (Thermo Fisher Scientific) or SNPtype assay (Fluidigm, San Francisco, CA, USA) according to the manufacturer's instructions. The PubMed database was used to find previously reported variants within the intron related to tacrolimus PK. Variants with an MAF higher than 1% in East Asian populations were selected. A total of ten variants in six genes, including *CYP3A4*, *CYP3A5*, *ABCB1*, *NR1I2*, *SLCO1B3*, and *SUMO4*, were genotyped using the SNaPShot assay (Supplementary Table S11), while a total of thirteen variants in nine genes, including *ABCB1*, *ABCC2*, *ABCG2*, *CYP3A5*, *NR1I2*, *POR*, *PPARA*, *PPARD*, and *SUMO4*, were genotyped using the SNPtype assay (Supplementary Table S12). NGS sequencing results were confirmed by a SNaPshot assay for all of the samples at *CYP3A7* rs12360 and *CYP3A5* rs15524. Seven variants, chosen based on their potential gene functions, including *CYP3A7*, and associated with tacrolimus C_0 in the discovery set were further confirmed by a SNaPshot assay in the replication set.

Statistical analyses.   The tacrolimus C_0/D was calculated by dividing the tacrolimus C_0 by the corresponding daily dose. The log-transformed tacrolimus C_0/D was used for further analysis. Linear regression and analysis of variance (ANOVA) were used to test for the association between C_0/D and each polymorphism for each time period according to additive models. After false discovery rate (FDR) correction, a value of $P < 0.05$ was considered statistically significant[50]. The Hardy-Weinberg equilibrium was tested for each SNP using the chi-square test to compare the observed and expected genotype frequencies. SNPs were excluded based on divergence from the Hardy-Weinberg equilibrium ($P > 0.001$). Continuous or categorical variables at the baseline characteristics between the discovery and replication sets were analyzed with Student's *t*-test and the chi-square test or Fisher's exact test, respectively. LMM analysis was performed to identify variants and clinical covariates associated with the tacrolimus C_0/D at different time points in all discovery set, replication set, and combined set. An LMM was constructed to include genotypes and clinical covariates that affected the repeated measures of C_0/D in a univariate analysis with a significance value of $P < 0.20$. Subsequently, a second LMM was performed with selected covariates with a significance level of $P < 0.05$ for all retrieved variables in the full model. Statistical analysis was performed using R software (version 3.3.2, www.r-project.org). Haplotypes and haplotype frequencies were calculated using the Haploview software (v4.2, Massachusetts Institute of Technology, Cambridge, MA, USA).

References

1. Yeung, S. *et al.* Primary immunosuppression with tacrolimus and low-dose mycophenolate mofetil in renal transplant recipients. *Transplant Proc* **36**, 2084–2086, https://doi.org/10.1016/j.transproceed.2004.08.110 (2004).
2. Kim, I. W. *et al.* Identification of factors affecting tacrolimus level and 5-year clinical outcome in kidney transplant patients. *Basic Clin Pharmacol Toxicol* **111**, 217–223, https://doi.org/10.1111/j.1742-7843.2012.00892.x (2012).
3. Lampen, A. *et al.* Metabolism of the immunosuppressant tacrolimus in the small intestine: cytochrome P450, drug interactions, and interindividual variability. *Drug Metab Dispos* **23**, 1315–1324 (1995).
4. Iwasaki, K. Metabolism of tacrolimus (FK506) and recent topics in clinical pharmacokinetics. *Drug Metab Pharmacokinet* **22**, 328–335 (2007).
5. Passey, C. *et al.* Dosing equation for tacrolimus using genetic variants and clinical factors. *Br J Clin Pharmacol* **72**, 948–957, https://doi.org/10.1111/j.1365-2125.2011.04039.x (2011).
6. Zhang, H. J., Li, D. Y., Zhu, H. J., Fang, Y. & Liu, T. S. Tacrolimus population pharmacokinetics according to CYP3A5 genotype and clinical factors in Chinese adult kidney transplant recipients. *J Clin Pharm Ther* **42**, 425–432, https://doi.org/10.1111/jcpt.12523 (2017).

7. Tang, J. T. *et al*. Pharmacogenetic aspects of the use of tacrolimus in renal transplantation: recent developments and ethnic considerations. *Expert Opin Drug Metab Toxicol* **12**, 555–565, https://doi.org/10.1517/17425255.2016.1170808 (2016).

8. Birdwell, K. A. *et al*. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clin Pharmacol Ther* **98**, 19–24, https://doi.org/10.1002/cpt.113 (2015).

9. Andreu, F. *et al*. A New CYP3A5*3 and CYP3A4*22 Cluster Influencing Tacrolimus Target Concentrations: A Population Approach. *Clin Pharmacokinet* **56**, 963–975, https://doi.org/10.1007/s40262-016-0491-3 (2017).

10. Li, J. L. *et al*. Interactive effects of CYP3A4, CYP3A5, MDR1 and NR1I2 polymorphisms on tacrolimus trough concentrations in early postrenal transplant recipients. *Pharmacogenomics* **16**, 1355–1365. https://doi.org/10.2217/pgs.1315.1378, Epub2015 Jul 1331 (2015).

11. Liu, S. *et al*. The POR rs1057868-rs2868177 GC-GT diplotype is associated with high tacrolimus concentrations in early post-renal transplant recipients. *Acta Pharmacol Sin* **37**, 1251–1258, https://doi.org/10.1038/aps.2016.77 (2016).

12. Choi, Y. *et al*. A pharmacogenomic study on the pharmacokinetics of tacrolimus in healthy subjects using the DMETTM Plus platform. *Pharmacogenomics J* **17**, 105–106, https://doi.org/10.1038/tpj.2016.85 (2017).

13. Liu, X. M. *et al*. [Associations of SUMO4 polymorphisms with tacrolimus concentrations in Chinese renal transplant recipients]. *Yao Xue Xue Bao* **50**, 180–184 (2015).

14. Do, R., Kathiresan, S. & Abecasis, G. R. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet* **21**, R1–9, https://doi.org/10.1093/hmg/dds387 (2012).

15. Gulati, A. & Somlo, S. Whole exome sequencing: a state-of-the-art approach for defining (and exploring!) genetic landscapes in pediatric nephrology. *Pediatr Nephrol.* https://doi.org/10.1007/s00467-017-3698-0 (2017).

16. Stevens, J. C. *et al*. Developmental expression of the major human hepatic CYP3A enzymes. *J Pharmacol Exp Ther* **307**, 573–582, https://doi.org/10.1124/jpet.103.054841 (2003).

17. Sim, S. C., Edwards, R. J., Boobis, A. R. & Ingelman-Sundberg, M. CYP3A7 protein expression is high in a fraction of adult human livers and partially associated with the CYP3A7*1C allele. *Pharmacogenet Genomics* **15**, 625–631 (2005).

18. Rodriguez-Antona, C., Jande, M., Rane, A. & Ingelman-Sundberg, M. Identification and phenotype characterization of two CYP3A haplotypes causing different enzymatic capacity in fetal livers. *Clin Pharmacol Ther* **77**, 259–270 (2005).

19. Williams, J. A. *et al*. Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug Metab Dispos* **30**, 883–891 (2002).

20. Kamdem, L. K. *et al*. Contribution of CYP3A5 to the *in vitro* hepatic clearance of tacrolimus. *Clin Chem* **51**, 1374–1381, https://doi.org/10.1373/clinchem.2005.050047 (2005).

21. Betts, S., Bjorkhem-Bergman, L., Rane, A. & Ekstrom, L. Expression of CYP3A4 and CYP3A7 in Human Foetal Tissues and its Correlation with Nuclear Receptors. *Basic Clin Pharmacol Toxicol* **117**, 261–266, https://doi.org/10.1111/bcpt.12392 (2015).

22. Crettol, S. *et al*. CYP3A7, CYP3A5, CYP3A4, and ABCB1 genetic polymorphisms, cyclosporine concentration, and dose requirement in transplant recipients. *Ther Drug Monit* **30**, 689–699, https://doi.org/10.1097/FTD.0b013e31818a2a60 (2008).

23. Elens, L. *et al*. 1199G>A and 2677G>T/A polymorphisms of ABCB1 independently affect tacrolimus concentration in hepatic tissue after liver transplantation. *Pharmacogenet Genomics* **17**, 873–883, https://doi.org/10.1097/FPC.0b013e3282e9a533 (2007).

24. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).

25. Lee, S. S. *et al*. Identification of a null allele of cytochrome P450 3A7: CYP3A7 polymorphism in a Korean population. *Mol Biol Rep* **37**, 213–217, https://doi.org/10.1007/s11033-009-9608-1 (2010).

26. Genvigir, F. D. *et al*. Influence of the CYP3A4/5 genetic score and ABCB1 polymorphisms on tacrolimus exposure and renal function in Brazilian kidney transplant patients. *Pharmacogenet Genomics* **26**, 462–472, https://doi.org/10.1097/FPC.0000000000000237 (2016).

27. Provenzani, A. *et al*. Influence of CYP3A5 and ABCB1 gene polymorphisms and other factors on tacrolimus dosing in Caucasian liver and kidney transplant patients. *Int J Mol Med* **28**, 1093–1102, https://doi.org/10.3892/ijmm.2011.794 (2011).

28. Hebert, M. F. *et al*. Interpreting tacrolimus concentrations during pregnancy and postpartum. *Transplantation* **95**, 908–915, https://doi.org/10.1097/TP.0b013e318278d367 (2013).

29. van Duijnhoven, E. M. *et al*. Increase in tacrolimus trough levels after steroid withdrawal. *Transpl Int* **16**, 721–725, https://doi.org/10.1007/s00147-003-0615-1 (2003).

30. Thompson, E. E., Kuttab-Boulos, H., Yang, L., Roe, B. A. & Di Rienzo, A. Sequence diversity and haplotype structure at the human CYP3A cluster. *Pharmacogenomics J* **6**, 105–114, https://doi.org/10.1038/sj.tpj.6500347 (2006).

31. Birdwell, K. A. *et al*. The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenet Genomics* **22**, 32–42, https://doi.org/10.1097/FPC.0b013e32834e1641 (2012).

32. Liu, M. Z. *et al*. IL-3 and CTLA4 gene polymorphisms may influence the tacrolimus dose requirement in Chinese kidney transplant recipients. *Acta Pharmacol Sin* **38**, 415–423, https://doi.org/10.1038/aps.2016.153 (2017).

33. Tamashiro, E. Y. *et al*. Influence of CYP3A4 and CYP3A5 polymorphisms on tacrolimus and sirolimus exposure in stable kidney transplant recipients. *Drug Metab Pers Ther* **32**, 89–95, https://doi.org/10.1515/dmpt-2016-0036 (2017).

34. Hesselink, D. A. *et al*. Genetic polymorphisms of the CYP3A4, CYP3A5, and MDR-1 genes and pharmacokinetics of the calcineurin inhibitors cyclosporine and tacrolimus. *Clin Pharmacol Ther* **74**, 245–254, https://doi.org/10.1016/S0009-9236(03)00168-1 (2003).

35. Sherry, S. T., Ward, M. & Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome research* **9**, 677–679 (1999).

36. Kiezun, A. *et al*. Exome sequencing and the genetic basis of complex traits. *Nat Genet* **44**, 623–630, https://doi.org/10.1038/ng.2303 (2012).

37. Cohen, T., Levin, L. & Mishmar, D. Ancient Out-of-Africa Mitochondrial DNA Variants Associate with Distinct Mitochondrial Gene Expression Patterns. *PLoS Genet* **12**, e1006407, https://doi.org/10.1371/journal.pgen.1006407 (2016).

38. Beck, T., Hastings, R. K., Gollapudi, S., Free, R. C. & Brookes, A. J. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur J Hum Genet* **22**, 949–952, https://doi.org/10.1038/ejhg.2013.274 (2014).

39. Bruno, A. E. *et al*. miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3′UTRs of human genes. *BMC Genomics* **13**, 44, https://doi.org/10.1186/1471-2164-13-44 (2012).

40. Oetting, W. S. *et al*. Genomewide Association Study of Tacrolimus Concentrations in African American Kidney Transplant Recipients Identifies Multiple CYP3A5 Alleles. *American journal of transplantation: official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **16**, 574–582, https://doi.org/10.1111/ajt.13495 (2016).

41. Loman, N. J. *et al*. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434–439, https://doi.org/10.1038/nbt.2198 (2012).

42. Sandmann, S. *et al*. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci Rep* **7**, 43169, https://doi.org/10.1038/srep43169 (2017).

43. Zhang, G. *et al*. Comparison and evaluation of two exome capture kits and sequencing platforms for variant calling. *BMC Genomics* **16**, 581, https://doi.org/10.1186/s12864-015-1796-6 (2015).

44. Damiati, E., Borsani, G. & Giacopuzzi, E. Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet* **135**, 499–511, https://doi.org/10.1007/s00439-016-1656-8 (2016).

45. Li, C. I., Samuels, D. C., Zhao, Y. Y., Shyr, Y. & Guo, Y. Power and sample size calculations for high-throughput sequencing-based experiments. *Brief Bioinform.* https://doi.org/10.1093/bib/bbx061 (2017).

46. Purcell, S., Cherny, S. S. & Sham, P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
47. World Medical, A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–2194, https://doi.org/10.1001/jama.2013.281053 (2013).
48. Min, S. I. *et al.* CYP3A5 *1 allele: impacts on early acute rejection and graft function in tacrolimus-based renal transplant recipients. *Transplantation* **90**, 1394–1400, https://doi.org/10.1097/TP.0b013e3181fa93a4 (2010).
49. Li, M. J. & Wang, J. Current trend of annotating single nucleotide variation in humans–A case study on SNVrap. *Methods* **79-80**, 32–40, https://doi.org/10.1016/j.ymeth.2014.10.003 (2015).
50. Yekutieli, Y. Ba. D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* **29**, 1165–1188 (2001).

Acknowledgements

Author Contributions

The study was designed by M.S., N.H., M.G.K., I.W.K. and J.M.O. Samples were collected by M.S., E.Y.S., S.I.M., and J.H. Genomic data were analyzed by H.S.J. and Y.K.S. Bioinformatics was performed by M.S., J.G., and I.W.K. The manuscript was prepared by M.S., I.W.K., and J.M.O. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-36085-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.