Genome-Wide Association Scan of Dupuytren's Disease

Joshua O. Ojwang, PhD, Indra Adrianto, PhD, Courtney Gray-McGuire, PhD, Swapan K. Nath, PhD, Celi Sun, MS, Kenneth M. Kaufman, PhD, John B. Harley, MD, PhD, Ghazi M. Rayan, MD

Purpose Dupuytren's disease (DD) has a strong genetic component that is suggested by population studies and family clustering. Genetic studies have yet to identify the gene(s) involved in DD. The purpose of this study was to identify regions of the entire genome (chromosomes 1–23) associated with the disease by performing a genome-wide association scan on DD patients and controls.

Methods We isolated genomic DNA from saliva collected from 40 unrelated DD patients and 40 unaffected controls. We conducted the genotyping using CytoSNP-Infinium HD Ultra genotyping assay on the Illumina platform. Using both log regression and mapping by admixture linkage disequilibrium analysis methods, we analyzed the single nucleotide polymorphism genotyping data.

Results Single nucleotide polymorphism analysis revealed a significant association in regions for chromosomes 1, 3 through 6, 11, 16, 17, and 23. Mapping by admixture linkage disequilibrium analysis showed ancestry-associated regions in chromosomes 2, 6, 8, 11, 16, and 20, which may harbor DD susceptibility genes. Both analysis methods revealed loci association in chromosomes 6, 11, and 16.

Conclusions Our data suggest that chromosomes 6, 11, and 16 may contain the genes for DD and that multiple genes may be involved in DD. Future genetic studies on DD should focus on these areas of the genome. *(J Hand Surg 2010;35A:2039–2045. © 2010 Published by Elsevier Inc. on behalf of the American Society for Surgery of the Hand.)*

Key words Dupuytren's disease, Dupuytren's disease genetics, admixture mapping, SNP analysis.

DUPUYTREN'S DISEASE (DD) is the most common heritable disorder of connective tissue, but sporadic cases may be encountered. The disease is characterized by progressive fibroblastic proliferation of components of the palmar fascial complex leading to digital contracture, most frequently affecting the ring and small fingers. In advanced stages of the disease the flexion deformities interfere with hand function and the ability to grasp and manipulate objects.[1] Palmar fascial proliferation may occur in non-Dupuytren's disease,[2] a clinical entity that should not be confused with Dupuytren's disease, in which palmar fascial proliferation usually follows trauma or surgery to the hand. Patients affected by palmar fascial proliferation of non-Dupuytren's disease can be of any age, gender, or race and may be diabetic with no family history of DD. The condition is unilateral and nonprogressive, usually only one hand is affected, and there is minimal digital involvement but no flexion deformity.

Typically, DD affects men of Northern European heritage, with a peak incidence at around 50 years of age. The condition is usually bilateral with progressive digital contracture at various rates. More than one digit is usually involved, and patients may have ectopic dis-

From the Department of Arthritis and Immunology, Oklahoma Medical Research Foundation; the Orthopedic Surgery Department, Oklahoma University and Division of Hand Surgery, Integris Baptists Medical Center; the Department of Medicine, University of Oklahoma Health Sciences Center; and the US Department of Veterans Affairs Medical Center, Oklahoma City, OK.

The authors thank Kim L. Nguyen for the preparation of DNA, Adam Adler, and Mai Li Zhu for help with genotyping, and Celi Sun for data analysis.

Received for publication May 24, 2010; accepted in revised form August 9, 2010.

Supported in part by National Institutes of Health grants NO1-AR62277, PO1-AI 083194, R37-AI 24717, RO1-AR42460, PO1-AR049084, P20-RR020143, R01 AI045050, and P30 AR053483; the US Department of Veteran Affairs; the Alliance for Lupus Research; and the Lupus Family Registry and Repository.

No benefits in any form have been received or will be received related directly or indirectly to the subject of this article.

Corresponding author: Ghazi M. Rayan, MD, Hand Microsurgery, 3366 NW Expressway, Oklahoma City, OK 73112; e-mail: ouhsgmr@aol.com.

0363-5023/10/35A12-0019$36.00/0
doi:10.1016/j.jhsa.2010.08.008

ease.[3] Onset at a younger age is associated with a more aggressive disease course and an increased risk for recurrence after treatment.[4,5] Prevalence is highest in elderly men from Scotland, Norway, and Iceland, and can be as high as 40%, but DD was reported among all ethnic groups at lower prevalence rates[6] including black African[7] and Japanese.[8] There is a higher disease prevalence in men, with a male-to-female ratio of approximately 6 to 1,[5,9,10] but with advancing age the incidence among women increases.[11] Nevertheless, the disease can be milder in women.[12]

Dupuytren's disease is a familial disorder with a strong genetic predisposition; variable autosomal dominance is the most likely pattern of inheritance. DD can be influenced by environmental factors including alcoholism, diabetes, and smoking.[5,10,13]

Numerous treatment options have been used, both surgical and nonsurgical. These treatment modalities are effective for controlling but not curing the disease. Understanding the molecular pathogenesis of DD is necessary for the development of new, more curative therapeutic alternatives.

No single gene responsible for the development of DD has been identified thus far, which suggests that DD may have a complex multifactorial (conditions arising from a combination of environmental and multiple genetic factors) etiology. Complex disorders such as systemic lupus erythematosus, diabetes, and certain cancers result from the combined action of environmental factors and alleles of more than one gene. The inheritance pattern of such disorders is usually complex compared with monogenic disorders and depends on the simultaneous influence of multiple alleles. DD is the most common hereditary connective tissue disorder among Caucasians[14]; hence, locating gene(s) for this disease is of great importance because their identification would provide insight into the fundamental pathogenesis of the disease and suggest targets for prevention or medical intervention. Investigating the genetic nature of DD by localizing regions that may harbor DD susceptibility genes will aid in designing more intricate genetic studies that may identify the precise gene(s) and causative variants within these gene(s). The purpose of this study was to identify regions that may harbor DD susceptibility genes by scanning the entire genome in a group of DD patients.

MATERIALS AND METHODS

Study subjects

We selected 2 groups for this study: DD patients and controls. All subjects of both groups were Caucasians of European ancestry. We identified DD patients by the presence of minimal phenotypic characteristics of the condition, including unequivocal palmar and digital cords, bilateral disease, multiple digits affected, and progressive digital contractures. The control group included volunteers who had no family history of DD; examination of their hands was normal with no evidence of palmar cutaneous thickening, nodules, or cords. We obtained genomic DNA from saliva samples from 40 unrelated DD patients (26 men and 14 women) and 40 unaffected controls (34 men and 6 women). Both INTEGRIS Baptist Medical Center and Oklahoma Medical Research Foundation institutional review boards approved the study and we obtained all samples with the written informed consent of subjects.

TABLE 1. Clinical Features of Subjects

Characteristic	DD-Affected Patients (n = 40)	Unaffected Controls (n = 40)
Gender (male/female)	26/14	34/6
Ethnicity	Caucasian	Caucasian
Age at sample (mean ± SD)	60.71 ± 11.15	40 ± 20
Patients with family history	20/20	0/40

We conducted genotype experiments to determine DD genetic association for 80 participants, all who resided in the same geographic area (Oklahoma City). Table 1 lists the clinical characteristics of both groups. Half of the patients in this group had a family history of the disease. The ages of unaffected control participants ranged from 20 to 60 years.

Genomic DNA extraction

We collected genomic DNA using an Oragene-DNA kit (DNA Genotek, Inc., Ontario, Canada). The Oragene-DNA immediately stabilizes DNA in the saliva upon mixing, and the collected Oragene-DNA/saliva samples are stable at room temperature for years without processing. Samples were processed no later than 4 months after collection. We performed saliva collection and genomic DNA extraction according to the manufacturer's instructions. The extracted DNA was evaluated for quality using a NanoDrop ND-1000 Spectrophotometer (Fisher/Thermo Scientific, Pittsburgh, PA). We calculated the ratios at 260/280 and 260/230, as well as the concentration of genomic DNA. The accepted ratios for 260/280 (protein) and 260/230 (organic material, phenol, etc.) are between 1.5 and 2.0 and 0.0 and 3.0, respectively. For each sample, the median yield of genomic DNA from 2 mL of saliva captured in 2 mL of Oragene-DNA was 100 mg. Ex-

tracted DNA from the unrelated DD patients and unaffected controls was stored at –20°C until it was needed for genotyping.

Genotyping

We genotyped samples using the Illumina HumanCyto SNP-12 array employing Infinium HD Assay Ultra genotyping assay methods (~300,000 single nucleotide polymorphisms [SNPs]) according to the manufacturer's instructions at the genotyping facilities located at the Oklahoma Medical Research Foundation (Oklahoma City, OK). More information on Illumina genotyping can be found at http://www.illumina.com/. We used genotype data only from samples with a call rate (the number of SNPs receiving a genotype call "AA", "AB", or "BB" divided by the total number of SNPs for each sample)[15] greater than 90% of the SNPs screened. The average call rate for all samples was 97.18%.

Quality control

Quality control is a necessary step to avoid false-positive or false-negative results from our statistical analyses resulting from poor-quality DNA samples or genotyping errors. We performed quality control for the statistical analysis methods described below as follows: for single SNP analysis, we assessed the quality of the genotype for each tested SNP by predetermined quality control inclusion criteria: minor allele frequency greater than 1%, SNP call rate greater than 90%, individual genotyping rate greater than 90%, and Hardy–Weinberg equilibrium (HWE) ($p>.001$) among all samples. In the mapping by admixture linkage disequilibrium (MALD) analysis, we performed data quality control by including individuals with a genotyping rate greater than 90% and SNPs with a call rate greater than 95%, minor allele frequency less than 1%, and HWE of $p>.001$.[16] HWE is the equilibrium state of a locus in which both allele and genotype frequencies in a population remain constant under appropriate conditions including random mating, no migration, no inbreeding, no mutation, no natural selection, and large population size.[17] Deviations from HWE can be due to genotyping error, chance, assumption violations, or a gene–disease association.[18]

Statistical analysis

Single SNP analysis: We calculated allele and genotype frequencies for each locus and tested them for HWE in controls. Case-control association studies were analyzed by chi-square test using 2×3 and 2×2 contingency tables of genotype and allele frequencies, respectively. We calculated odds ratios and p values using PLINK,[19] a free, open-source whole genome association analysis toolset designed to perform a range of basic, large-scale analyses in a computationally efficient manner. A p value explains the strength of association between an SNP and a disease; $p<1 \times 10^{-4}$ is considered statistically significant.

MALD analysis: We used the MALD method with the ADMIXMAP program/software[20,21] to localize disease-causing genetic variants that differ in frequency across different ancestral populations. Differences in the proportion of admixture for a particular chromosomal segment between cases and controls can indicate that a region is involved in a disease. Case-only analysis can also be done by looking for differences in admixture proportions between specific regions and the rest of the genome in the same individual. This analysis method uses Bayesian and classic approaches to perform admixture-mapping analyses. Markov Chain Monte Carlo simulation[22] is used to calculate the distribution of all unobserved variables given the observed genotypes, trait data, and prior ancestral allele frequencies. These unobserved variables include the ancestry at each locus and the ancestry-specific allele frequencies at each locus.[17] ADMIXMAP compares observed versus expected ancestry across the genome. The readout for the method is a Z-score, which is a statistical test for association with ancestry at each locus, comparing the observed and expected proportions of gene copies at each locus. The Z-score is considered significant at $|z\ score|$ greater than 3.

To complete the ADMIXMAP analysis, we provided founder (Northern European) and nonfounder (Southern European) population allele frequencies using publicly available control data from the Illumina iControlDB Hap550v1 and Hap550v3 (http://www.illumina.com/science/icontroldb.ilmn). We clustered Northern and Southern European samples and conducted an association analysis to identify SNPs that can distinguish those samples. We identified 3,133 SNPs that are informative to separate Northern and Southern European samples. These SNPs are called ancestry-informative markers when they have large allele frequency differences between those populations. We then calculated Northern and Southern European allele frequencies of those SNPs/ancestry-informative markers based on 432 Northern European and 121 Southern European samples using publicly available control data from the Illumina iControlDB Hap550v1 and Hap550v3 (http://www.illumina.com/science/icontroldb.ilmn). We hypothesized that we could localize dis-

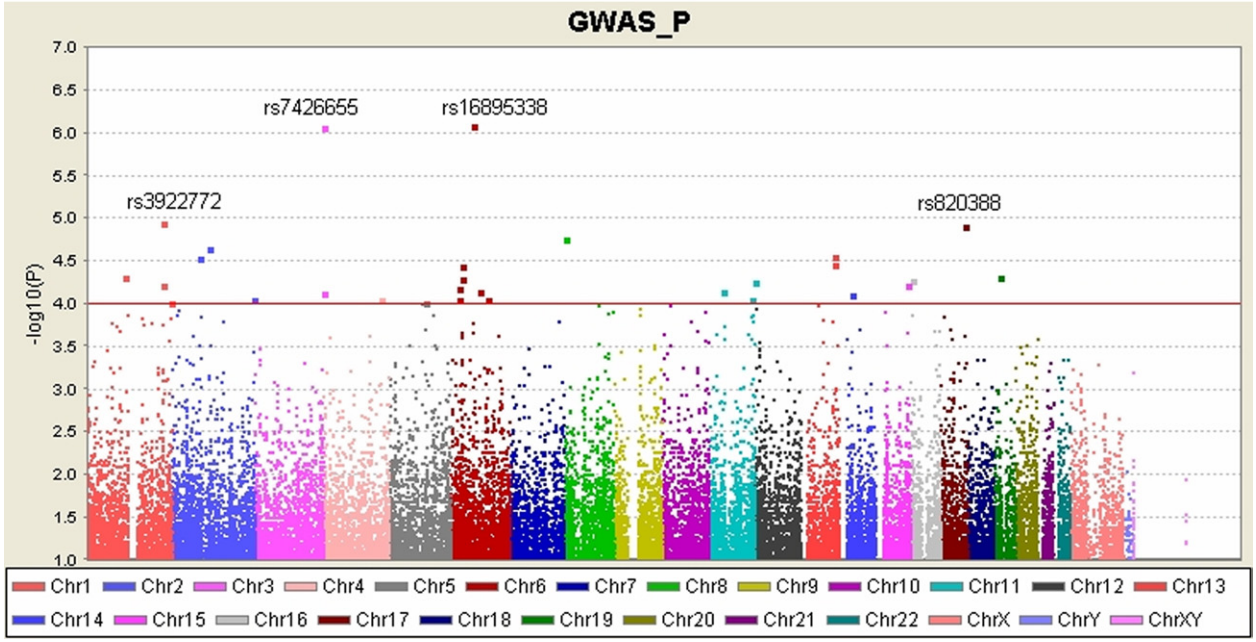

FIGURE 1: Location of SNPs (each dot) in all 22 and 3 gender chromosomes on the x axis and p values of allelic association on the y axis. SNPs with significance of association are those above the red line ($p<1 \times 10^{-4}$).

ease-causing genetic variants by comparing the allele frequencies between those populations.[17]

The DD dataset is from 80 samples, 40 cases and 40 controls with approximately 300,000 SNPs. We then selected 3,133 ancestry-informative markers from this dataset for admixture mapping. We performed standard data quality control, which resulted in one control with greater than 10% missing genotypes to be removed. To increase the number of controls, we added 2 publicly available control sets: Illumina iControls (432 Northern Europeans and 121 Southern Europeans) and HapMap CEU (60 samples). HapMap CEU is the International HapMap Project sample collection of Utah residents with Northern and Western European ancestry taken from the Centre d'Etude du Polymorphisme Humain collection. Then, we performed ADMIXMAP analyses on the Dupuytren's data only and Dupuytren's plus iControl plus CEU data for both case-only and case-control analyses across the entire genome (chromosomes 1–23).

RESULTS

Single SNP association

To evaluate the variation as a result of DD phenotype in patients, we performed the SNP analysis on 40 unrelated cases and 40 unaffected controls. The participants showed no problematic population stratification; however, we identified 3 individuals (affected) as outliers when we conducted principal component evaluation to correct for population stratification.[23] We removed 4 individuals (controls) because of the low genotyping (greater than 10% missing genotypes for each individual); 37 affected and 36 controls remained for the analysis. There was a total of 301,232 SNPs, which was reduced to 251,837 after we performed quality control.

Figure 1 shows the genome-wide association scan output. The y axis shows the p values (significance of association) and each dot represents an SNP evaluated in the study. Any SNP above the red line is considered significant ($p<1 \times 10^{-4}$) because this indicates that the association around that SNP is less likely owing to chance. The x axis shows the SNP locations on the 25 chromosomes.

Table 2 lists the results of the most significant SNPs and nearest genes. The associated SNPs are located in chromosomes 1, 3 through 6, 11, 16, 17, and 23.

MALD association

Because DD is mostly prevalent in the Northern European population, we took advantage of ADMIXMAP/MALD analysis method to evaluate the DD genetic association of the genotype data based on ancestry. To accomplish this task, we used the Illumina iControl data to determine Northern European and Southern European population allele frequencies. Results from ADMIXMAP analysis as shown in Figure 2 indicated that ancestry-associated regions with a z score

TABLE 2. Most Important SNPs, Along With Their Respective Chromosomes, Genes, or Nearest Genes, Before (Plink_P) and After Correcting Population Stratification (eigen_corrected_P)

CHR	SNP	BP	Nearby Gene	A1	F_A	F_U	A2	CHISQ	OR	L95	U95	Plink_P	Eigen_ Corrected_P
1	rs12032381	219683699	LOC100132626	C	0.361	0.083	A	16.07	6.22	2.37	16.31	6.10E-05	6.1085E-06
1	rs1903138	219719126	LOC100132626	C	0.378	0.111	A	14.03	4.87	2.02	11.65	0.00018	5.80846E-05
3	rs7426655	197829546	FBXO45‖LRRC33	C	0.014	0.314	A	24.25	0.03	0.004	0.23	8.48E-07	4.5602E-05
4	rs4864039	132525024	CYCSP14	G	0.625	0.319	A	13.49	3.55	1.79	7.06	0.00024	6.45995E-05
5	rs6897647	15957024	FBXL7	A	0.189	0.014	G	12.17	16.6	2.12	129.7	0.00049	7.97842E-05
6	rs3132506	31284205	HCG27‖HLA-C	G	0.392	0.097	A	17.06	5.98	2.41	14.85	3.63E-05	8.58496E-05
6	rs3130473	31307187	HCG27‖HLA-C	A	0.405	0.111	G	16.42	5.46	2.29	13.01	5.09E-05	5.31839E-05
6	rs16895338	65267311	LOC727945	A	0.455	0.076	G	24.3	10.2	3.62	28.55	8.24E-07	6.66237E-05
11	rs2846236	123571837	OR10D3P‖OR8F1P	G	0.284	0.556	A	11.08	0.32	0.16	0.63	0.00087	7.70025E-05
11	rs6590281	127224748	LOC100132514	A	0.081	0.333	G	14.22	0.18	0.07	0.46	0.00016	8.43636E-05
16	rs1919060	5258657	LOC100129495	G	0.297	0.056	A	14.57	7.19	2.335	22.2	0.00014	6.49485E-05
16	rs11649669	5293566	LOC100131502	A	0.432	0.129	G	16.31	5.16	2.24	11.9	5.38E-05	9.76969E-05
17	rs1978136	29401811	ACCN1	C	0.162	0.444	A	13.81	0.24	0.11	0.52	0.00020	2.44881E-05
23	rs17335275	3584538	PRKX	G	0.204	0.524	A	10.14	0.2	0.093	0.59	0.00145	8.50178E-05

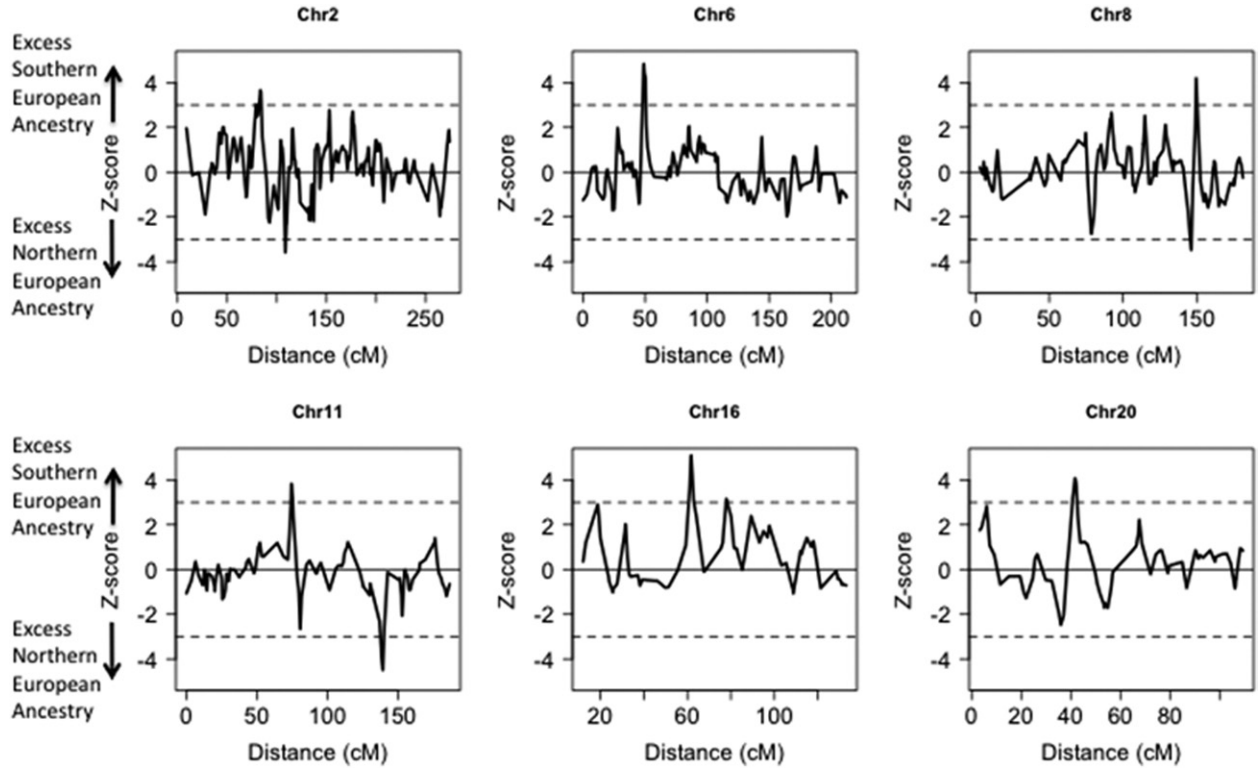

FIGURE 2: Charts of 6 chromosomes showing statistically significant z score values above or below the dotted lines at +3 or –3. The x axis is the genetic distance in centimorgans.

greater than 3 in chromosomes 2, 6, 8, 11, 16, and 20 may contain DD susceptibility genes.

We identified 3 chromosomes as common to both analysis methods: 6, 11, and 16. The ancestry association on chromosome 6 and single SNP association on the same chromosome are within 10,000 base pairs. These observations are notable because the SNPs that are in proximity (typically, 50 kb apart or closer) are more likely than those that lie farther apart to have alleles that travel together in a block when passed from parent to child. This phenomenon is termed *linkage disequilibrium*.[24]

DISCUSSION

Dupuytren's disease is believed to have a strong genetic component, and identifying the gene(s) for the disease has thus far eluded researchers. Identifying the gene(s) responsible for DD will elucidate the nature of the disease, identify at-risk individuals, distinguish between genetic and sporadic cases, validate the clinical observations about non-DD, and offer new methods of treatment and hope for cure rather than controlling the diseases with current treatment methods.

Our study provided strong evidence of the genetic nature of DD. Using a genome-wide association scan, we attempted to search for regions on the genome or genes that may have a major role in the pathogenesis of DD. Although our study did not identify the precise gene loci, data analysis by both ancestry and SNP-based methods revealed that regions on chromosomes 6, 11, and 16 might contain DD susceptibility genes. The region associated with chromosome 6 is near gene HCG27, which is located in the human leukocyte antigen locus. The finding about this region was intriguing because it was identified as associated with DD by both ancestry and SNP-based analysis methods. The most notable SNPs associated with DD in our study on chromosome 16p13 is 36 centimorgans (1 centimorgan corresponds to about 1 million base pairs in human beings on average) away from the main associated region on chromosome 16q previously reported on the Swedish family linkage analysis.[5] None of the proposed DD-associated polymorphisms in transforming growth factor-β receptor 1 and ZF9, reported by Bayat et al.,[9] were present in our study. The relatively small number of patients was a limitation of our study. However, that our data and others have identified several regions on multiple chromosomes to be associated or linked to DD strongly suggests that DD is an oligogenic disorder that results from the combined action of alleles of more than

one gene. This must be taken into consideration when pursuing future genetic studies on DD.

We obtained the data reported here from unrelated DD patients; half of these patients had no known family history of the disease. It is uncertain whether patients without family history are truly sporadic cases or are simply unaware of an existing family history. Genetic testing on a more homogeneous group of DD patients with known family history and strong diathesis, along with sampling of the proband and affected family members, may provide information about the precise location of the culprit genes.

The work reported here is encouraging because the analysis power of this small sample size based on the results is significant, and we believe that future experiments using additional unrelated DD patients and family samples will have enough power to unequivocally identify DD-associated genes. Future genetic studies should focus on areas of the genome that most likely contain the genes for DD in chromosomes 6, 11, and 16. DD is one of the few remaining autosomal-dominant diseases without known causative gene(s). We believe that the results of this work in combination with already published reports will greatly aid in designing more detailed genetic studies to identify the precise underlying genes and causative variants within the genes.

REFERENCES

1. Rayan GM. Dupuytren disease: anatomy, pathology, presentation, and treatment. J Bone Joint Surg 2007;89A:189–198.
2. Rayan GM, Moore J. Non-Dupuytren's disease of the palmar fascia. J Hand Surg 2005;30B:551–556.
3. McFarlane RM, Classen DA, Porte AM, Botz JS. The anatomy and treatment of camptodactyly of the small finger. J Hand Surg 1992; 17A:35–44.
4. Hueston JT. Unsatisfactory results in Dupuytren's contracture. Philosophies of Dr. J. T. Hueston. Hand Clin 1991;7:759–763.
5. Hu FZ, Nystrom A, Ahmed A, Palmquist M, Dopico R, Mossberg I, et al. Mapping of an autosomal dominant gene for Dupuytren's contracture to chromosome 16q in a Swedish family. Clin Genet 2005;68:424–429.
6. Saboeiro AP, Porkorny JJ, Shehadi SI, Virgo KS, Johnson FE. Racial distribution of Dupuytren's disease in Department of Veterans Affairs patients. Plast Reconstr Surg 2000;106:71–75.
7. Muguti GI, Appelt B. Dupuytren's contracture in black Zimbabweans. Cent Afr J Med 1993;39:129–132.
8. Abe Y, Rokkaku T, Ofuchi S, Tokunaga S, Takahashi K, Moriya H. Dupuytren's disease on the radial aspect of the hand: report on 135 hands in Japanese patients. J Hand Surg 2004;29B:359–362.
9. Bayat A, Watson JS, Stanley JK, Ferguson MW, Ollier WE. Genetic susceptibility to dupuytren disease: association of Zf9 transcription factor gene. Plast Reconstr Surg 2003;111:2133–2139.
10. Lee LC, Zhang AY, Chong AK, Pham H, Longaker MT, Chang J. Expression of a novel gene, MafB, in Dupuytren's disease. J Hand Surg 2006;31A:211–218.
11. Ross DC. Epidemiology of Dupuytren's disease. Hand Clin 1999; 15:53–62, vi.
12. Burge P. Genetics of Dupuytren's disease. Hand Clin 1999;15: 63–71.
13. Bayat A, Stanley JK, Watson JS, Ferguson MW, Ollier WE. Genetic susceptibility to Dupuytren's disease: transforming growth factor beta receptor (TGFbetaR) gene polymorphisms and Dupuytren's disease. Br J Plast Surg 2003;56:328–333.
14. Hindocha S, John S, Stanley JK, Watson SJ, Bayat A. The heritability of Dupuytren's disease: familial aggregation and its clinical significance. J Hand Surg 2006;31A:204–210.
15. Huentelman MJ, Craig DW, Shieh AD, Corneveaux JJ, Hu-Lince D, Pearson JV, et al. SNiPer: improved SNP genotype calling for Affymetrix 10K GeneChip microarray data. BMC Genomics 2005; 6:149.
16. Weale ME. Quality control for genome-wide association studies. Methods Mol Biol 2010;628:341–372.
17. Thomas DC. Statistical methods in genetic epidemiology. Oxford: Oxford University Press, 2004:435.
18. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:967–986.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
20. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, et al. Control of confounding of genetic associations in stratified populations. Am J Hum Genet 2003;72:1492–1504.
21. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. Am J Hum Genet 2004;74:965–978.
22. McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann Hum Genet 2000;64:171–186.
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
24. Christensen K, Murray JC. What genome-wide association studies can do for medicine. N Engl J Med 2007;356:1094–1097.