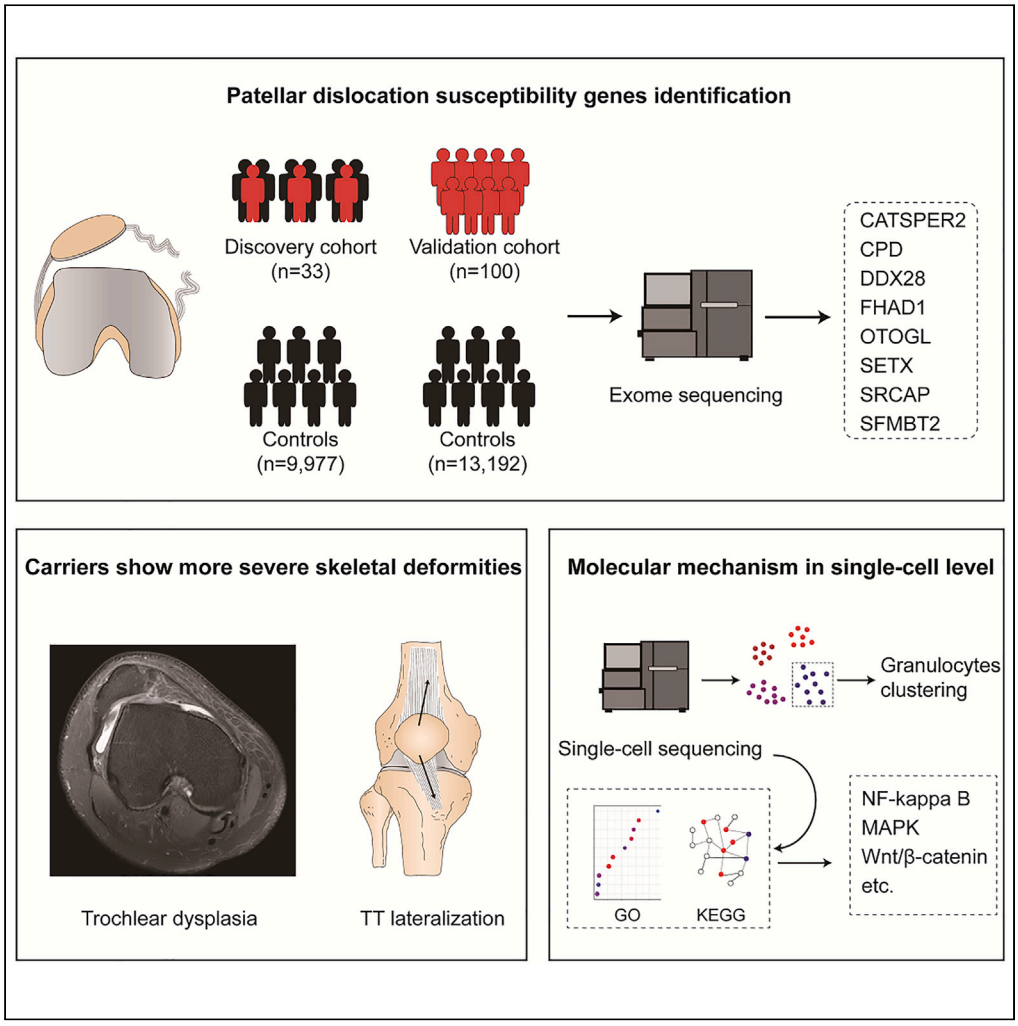
iScience

CellPress
OPEN ACCESS

Article

Identification of eight genes associated with recurrent patellar dislocation



Zijie Xu, Siyuan Huang, Yifan Song, ..., Haijun Wang, Ruibin Xi, Jia-Kuo Yu

bysywang@163.com (H.W.)
ruibinxi@math.pku.edu.cn (R.X.)
yujiakuo@126.com (J.-K.Y.)

Highlights

Exome sequencing identified eight patellar dislocation susceptible genes

Patients with identified disruptive variants show more severe skeletal deformities

The granulocyte is involved in the pathophysiological process of RPD

Xu et al., iScience 27, 109697
May 17, 2024
https://doi.org/10.1016/j.isci.2024.109697

iScience

CellPress
OPEN ACCESS

Article

Identification of eight genes associated with recurrent patellar dislocation

Zijie Xu,[2,3,9] Siyuan Huang,[4,9] Yifan Song,[1] Chao Xu,[5] Hongyu Yan,[6] Ouyang Linkun,[4] Bo Lv,[7] Fuzhen Yuan,[2,3] Bingbing Xu,[2,3] Haijun Wang,[2,3,*] Ruibin Xi,[8,*] and Jia-Kuo Yu[1,10,*]

SUMMARY

The inheritance of recurrent patellar dislocation (RPD) is known, but the susceptible gene remains unidentified. Here, we performed the first whole exome sequencing (WES) cohort study to identify the susceptible genes. The results showed eight genes were associated with this disease. Notably, the carboxypeptidase D (CPD) gene showed the highest relevance based on its gene function and tissue expression. Single-cell sequencing results indicate that the CPD gene is involved in the pathophysiological process of RPD through granulocytes. Implicated pathways include nuclear factor kappa B (NF-κB), mitogen-activated protein kinase (MAPK), and Wnt/β-catenin signaling, potentially influencing CPD's role in RPD pathogenesis. This study identified the susceptible gene and investigates the potential pathogenesis of RPD, which provided a new prospect for the understanding of RPD. Besides, it would offer the theoretical basis for disease prevention and genetic counseling.

INTRODUCTION

Recurrent patellar dislocation (RPD; MIM: 169000) is a knee joint disorder with an annual incidence of 77–108 per 100,000 persons.[1,2] It is characterized by the patella pathologically disarticulating out from the patellofemoral joint, which results in the fracture of cartilage and rupture of ligaments.[3,4] RPD is a well-known risk factor for osteoarthritis (OA) of the knee joint.[5] Therefore, understanding the etiology and pathogenesis of RPD is of paramount importance to the prevention of OA.

The observation of a high proportion of femoral trochlea dysplasia (>90%) and sex differences in incidence (female: male ratio, 2:1) suggests the possibility of heritability in RPD.[6] Additionally, RPD is a known symptom in several syndromic diseases, such as Down syndrome and Kabuki syndrome.[7] Autosomal inheritance of RPD has been reported in familial cases.[8] Several genes, such as TBC1D7 and ACAN, have been identified as potentially associated genes in case report studies.[9–11] However, none of these reported genes has been replicated in more than two families or a sporadic disease cohort. Immune responses and inflammation signaling have also been implicated in the pathophysiology of RPD.[12,13] Xu et al.[12] identified the involvement of JAK/STAT signaling in the pathogenesis of RPD through bulk RNA-seq, but single-cell RNA sequencing (scRNA-seq) has not yet been employed in RPD molecular mechanism investigation. Therefore, further studies are necessary to identify candidate genes and elucidate the molecular mechanism underlying RPD.

In this study, we conducted a study with a cohort of 533 RPD patients to investigate the inherited tendency and identify susceptible genes. Whole-exome sequencing (WES) was conducted on both familial and sporadic cases. A total of eight RPD-associated genes were identified. An increased mutation burden of these genes was detected in the RPD cohort. Patients with rare disruptive variants in these genes exhibited more severe skeletal deformity. Furthermore, we employed scRNA-seq on RPD patients to evaluate the specific expression pattern of the identified genes and reveal potential mechanisms at the single-cell resolution.

RESULTS

The basic characteristics of recurrent patellar dislocation

Between March 2019 and March 2021, 400 RPD patients were recruited prospectively to evaluate the clinical characteristics of RPD. The average onset age was 18.2 ± 7.3 years, with females comprising 66.5% of the cohort and 10.5% of them experiencing bilateral dislocation.

[1]Orthopaedic Sports Medicine Center, Beijing Tsinghua Changgung Hospital, Affiliated Hospital of Tsinghua University, Beijing, China
[2]Sports Medicine Department, Beijing Key Laboratory of Sports Injuries, Peking University Third Hospital, Beijing, China
[3]Institute of Sports Medicine, Peking University, Beijing, China
[4]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China
[5]The Department of Joint Surgery and Sports Medicine, The Second Affiliated Hospital of Xinjiang Medical University, Urumchi, China
[6]Department of Pediatric Neurology, Children's Hospital Affiliated to the Capital Institute of Pediatrics, Peking University, Beijing, China
[7]Department of Orthopedics, People's Hospital of Guilin, Guilin, Guangxi, China
[8]School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: bysywang@163.com (H.W.), ruibinxi@math.pku.edu.cn (R.X.), yujiakuo@126.com (J.-K.Y.)
https://doi.org/10.1016/j.isci.2024.109697

Table 1. Clinical features of recruited recurrent patellar dislocation patients

	FPD patients (n = 33)	SPD patients (n = 100)	p value
Age at onset, y	14.12 ± 4.24	16.68 ± 6.23	0.02
Sex, n			0.20
Male	6	33	
Female	27	67	
Affect side, n			0.41
Unilateral	26	85	
Bilateral	7	15	
Courses, n			0.19
Daily activity	23	36	
Sports activity	10	59	
Direct force	0	5	
Joint hypermobility, n	13	34	0.57
Trochlear dysplasia	33	90	0.06

Daily activities include walking, and squatting; the direct force includes slip and car accidence.
FPD, familial recurrent patellar dislocation; SPD, sporadic recurrent patellar dislocation.

Trochlear dysplasia was present in 92.5% of cases, while only 5.5% of cases were attributed to direct forces. A total of 34.3% of patients had joint hypermobility.[14] There were 10 cases in this cohort with a family history of RPD, and 90% of them were female. No cases had known chromosomal anomalies or molecularly diagnosed syndromes. Detailed patient phenotypes for this cohort are available in the supplement (Table S1). These results show gender disparity, skeletal and ligamentous deformities, and familial onset characteristics in RPD patients.

Familial RPD patients have an earlier onset age than the sporadic cases

The clinical data and peripheral blood samples of RPD patients were collected between April 2021 and May 2022. A total of 133 individuals with RPD were participants in this study, including 33 familial and 100 sporadic RPD cases. The clinical data and biological samples were collected for further analysis (Figure 1). Each familial case has at least one family member with diagnosed RPD. The familial cases showed an earlier onset age than the SPD cases (14.12 vs. 16.68, $p = 0.02$). Female patients predominated in both the FPD and SPD cases (67.0%–81.8%). Bilateral patellar dislocation was present in 15–21% of the included patients. All the FPD patients and 95% of SPD patients dislocated their patella without direct forces. Thirteen FPD patients (39.4%) and 34 SPD patients (34%) had joint hypermobility. Although the percentage of trochlear dysplasia was higher in FPD patients than in SPD patients, there was no statistical difference between the two groups (100% vs. 90%, $p = 0.06$). The detailed comparisons of clinical phenotypes were presented in Table 1. WES was performed in this cohort to identify candidate genes for RPD.

Enrichment analysis of rare genetic variants in RPD cohort

We investigated the biological processes and significant pathways that might associate with RPD. The sequenced data were processed according to the aforementioned criteria outlined in the STAR methods section. A total of 849 genes with a mutation frequency of more than 5% in RPD patients and less than 1% in the gnomAD database were identified. All these mutated genes were included in gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. The mutated genes were significantly enriched in GO terms such as cell adhesion, cell proliferation, and MAPK cascade. The results of KEGG analysis showed these genes were enriched in endocytosis and retinol metabolism pathways (Figure S1 in the Supplement). Alterations in cell adhesion and proliferation functions are closely associated with a variety of autoimmune diseases, such as OA.[15,16] RPD has clinical characteristics of bone deformities, and joint hypermobility, and is also a contributor to OA. The results of GO and KEGG analysis indicate abnormal autoimmune may involve in the potential mechanism of RPD.

The identification and prioritization of RPD-associated variants

We defined the familial cases as the discovery cohort, and to establish a suitable control group, we utilized the East Asia individuals from gnomAD v2.1.1 database[17] (n = 9,977). To identify the candidate genes of RPD, we employed the aforementioned criteria, based on allele frequency, and predicted impact on protein function. Our analysis detected ten variants (Figure 2A) that were enriched in the discovery cohort when compared to the general population (gnomAD-East Asia, $p < 0.05$). In order to validate the association between these variants and RPD, we compared their frequencies between 100 patients with sporadic RPD (validation cohort) and controls without RPD (ChinaMap[18] = 10,588, Gnomad v3.1.1-East Asian = 2,604). Among these ten variants, eight of them (CATSPER2, CPD, SETX, OTOGL, SRCAP, FHAD1, SFMBT2, and DDX28) were also significant in the validation cohort (Table 2; Table S2). Therefore, we considered these eight variants as candidates associated with an increased risk of RPD.

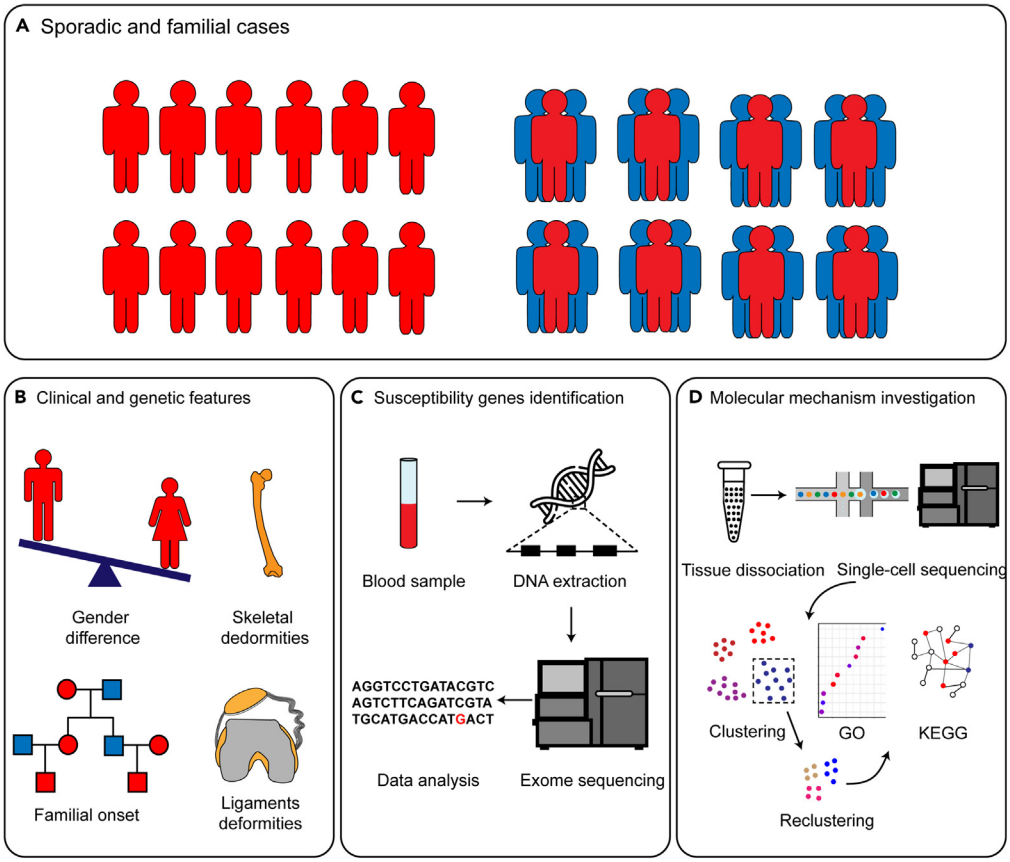


Figure 1. The study design and participants of this study

(A) The study participants of this study.

(B) The clinical data analysis of the participants.

(C) Whole exome sequencing to identify the susceptible genes of this disease.

(D) Single-cell sequencing to investigate the molecular mechanisms of this disease.

Through comparison of the frequency of the identified variants between familial and sporadic cases, we observed a higher proportion of familial cases carrying each of the eight variants, suggesting their potential role in the inheritance of RPD (Figure 2B). Notably, these missense mutations were predicted to have deleterious effects on protein function, as presented in Table S3, with the altered surface charge and structure of the proteins shown in Figure S2. To further explore the relevance of these candidate genes to RPD, we evaluated their expression in the three most relevant tissues, namely bone marrow, skeletal muscle, and cartilage. Our results showed that carboxypeptidase D (CPD) was highly expressed in all three tissues, indicating its potential involvement in RPD pathophysiological processes (Figures 2C–2E).

Increased mutation burden of RPD-associated genes in the disease cohort

A burden test of rare variants was conducted to evaluate the impact of the identified genes on RPD risk. LoF variants and missense mutations with a CADD score>15 were included in the analysis, and we compared the mutation burden of these genes in the RPD cohort of 133 cases to 141,565 controls from the gnomAD database (Table S4). Separate analyses were performed for the familial cohort, sporadic cohort, and total cohort. The results revealed that all eight genes exhibited a significantly increased mutation burden in the RPD cohort ($p < 0.001$, Figure 3; Table S6). Furthermore, the proportion of individuals carrying disruptive variants of the identified genes was higher in the familial RPD cohort, and the odds ratio of the burden test was greater than that of the sporadic cohort (Figures 3A and 3B). Notably, individuals with the identified disruptive variants had a higher risk of RPD incidence (Figure 3C).

More severe skeletal deformities in RPD patients with identified disruptive variants

To investigate which phenotype may be associated with RPD-associated genes, we divided the 133 RPD patients into the Mut group and non-Mut group based on whether they carried the variants of RPD-associated genes. Given that RPD patients often exhibit various skeletal and ligament-related phenotypes, we sought to determine whether specific phenotypes were associated with the RPD-associated genes.

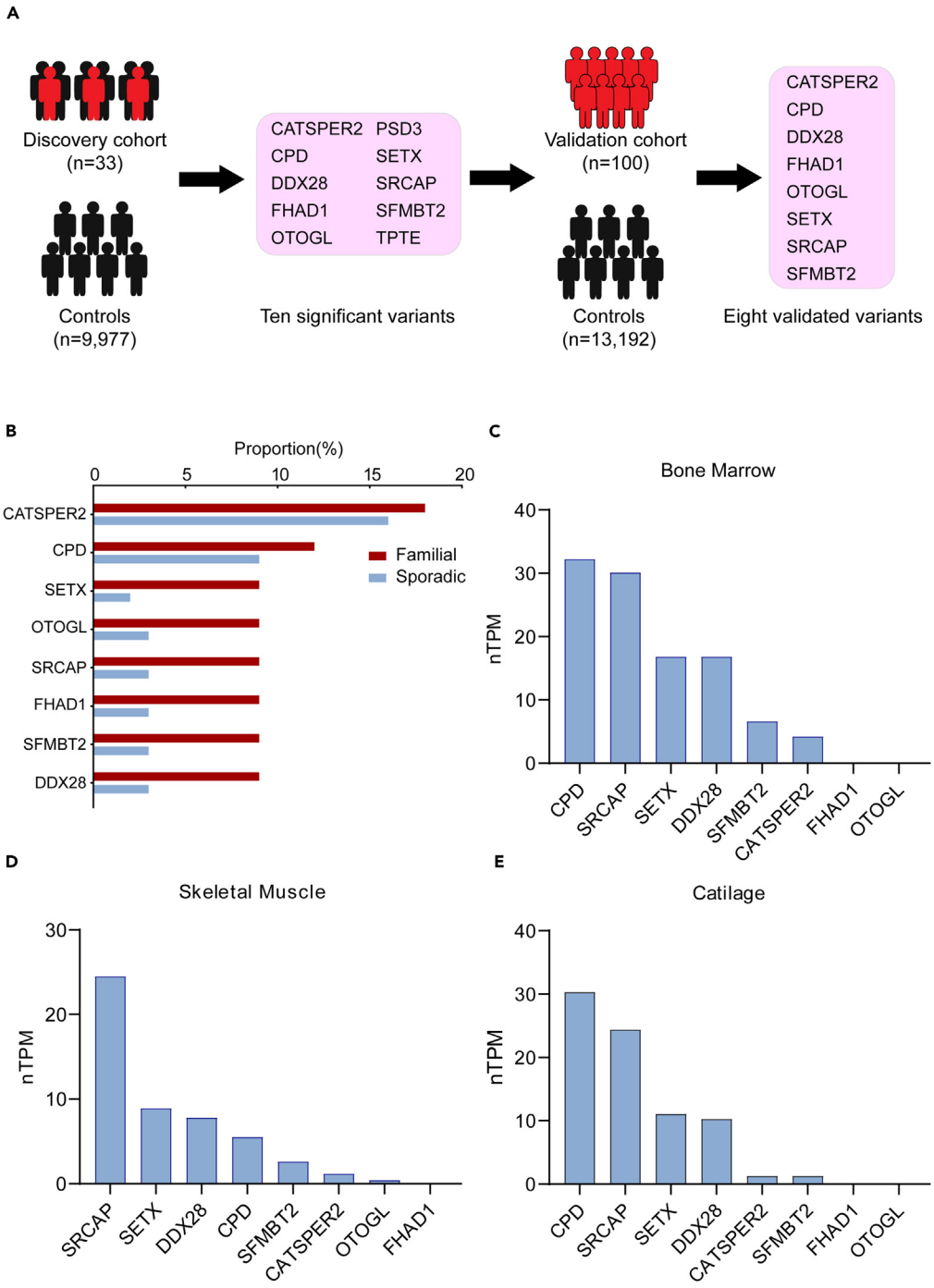


Figure 2. The identification and prioritization of RPD-associated genes

(A) The flowchart for susceptible gene identification.

(B) The proportion of cases carried the identified variants.

(C–E) The expression level of identified genes in bone marrow, skeletal muscle, and cartilage.

The onset age, gender distribution, and the proportion of joint hypermobility did not show significant differences between Mut and non-Mut groups (Figure S3 in the Supplement). However, we observed more severe skeletal deformities in the Mut group compared to the non-Mut group (Figure 4). Specifically, the Mut group exhibited a shallower trochlear groove compared to the non-Mut group (Figure 4A, $p < 0.001$), indicating more severe trochlear dysplasia. In addition, the Mut group displayed greater tibial tubercle lateralization (Figure 4B,

Table 2. Variants enriched in familial RPD and observed in sporadic RPD

		Discovery cohort vs. controls (33 vs. 9,977)			Validation cohort vs. controls (100 vs. 13,192)			Overall (133 vs. 23,169)		
Variant	Gene	Frequencies	OR	P	Frequencies	OR	P	Frequencies	OR	P
15:43632222:C:T	CATSPER2	6/63	31.53	1.04E-7	16/32	71.50	1.05E-22	22/95	43.83	3.83E-27
17:30379096:C:T	CPD	4/238	5.34	8.36E-3	9/550	2.21	0.04	13/788	2.97	7.83E-4
9:132330094:G:A	SETX	3/12	78.89	1.51E-5	2/9	29.58	2.9E-3	5/21	42.22	3.48E-7
12:80302754:A:T	OTOGL	3/30	31.58	1.74E-4	3/93	4.30	0.03	6/123	8.67	1.03E-4
16:30704076:G:A	SRCAP	3/28	33.87	1.44E-4	3/99	4.04	0.04	6/127	8.40	1.22E-4
1:15382087:T:C	FHAD1	3/27	35.09	1.30E-4	3/14	28.67	2.6E-4	6/41	26.07	2.88E-7
10:7171974:C:T	SFMBT2	3/9	105.21	7.37E-6	3/73	5.49	0.02	6/82	13.01	1.20E-5
16:68022758:C:A	DDX28	3/55	17.21	9.28E-4	4/126	4.25	0.02	7/181	6.89	1.12E-4

RPD, recurrent patellar dislocation.

$p < 0.007$), suggesting more severe tibia development and muscle trajectory abnormalities. Therefore, patients carrying the identified risk genes exhibited more severe clinical phenotypes, and the dysfunction of these genes may contribute to the onset of disease by affecting bone development.

We conducted a more detailed phenotypic analysis in patients carrying two or more mutated genes, and a total of 15 patients were included. None of these patients experienced patellar dislocation due to direct force, with most of the injuries resulting from daily activities. Cartilage damage was observed in all surgical patients, and 75% of these patients had severe cartilage damage (level>III). For patients who underwent bilateral knee radiological examinations, all of them had bilateral trochlear dysplasia, while 75% had bilateral patellar tilt and tibial tubercle lateralization (Table S6). A representative image of a patient with unilateral RPD was provided in Figure S4 in the Supplement.

Enrichment of CPD and other risk genes in human bone marrow

To gain insight into the biological role of the RPD-associated genes, we performed scRNA-seq analysis of bone marrow obtained from knee joints of three RPD patients and compared the results with the bone marrow scRNA-seq data from three non-RPD donors. We obtained 38,337 high-quality single cells which were clustered into 17 distinct cell clusters, including three T cell subsets (c0, c1, c4), two NK cell subsets (c2, c3), three granulocyte subsets (c5, c11, c15), three B cell subsets (c6, c7, c10), two osteoblast subset (c14, c16), one endotheliocyte subset (c8), one macrophage subset (c9), one erythroblast subset (c12), and one dendritic cell subset (c13) (Figure S5 in the Supplement).

We used marker genes to identify a total of 9 cell types (Figure 5A; Table S7; Figure S5B). CPD was significantly enriched in c5 ($p = 0$) and c11 ($p = 2.12 \times 10^{-31}$), which were identified as granulocyte (Figures 5A; Table S7). The enrichment of other RPD-associated genes was displayed in Figure S6. The expression of CPD was highest in granulocytes (Figure 5B), and the proportion of cells expressing CPD was highest in granulocytes compared to the other cell types (Figure 5C).

To further investigate the potential involvement of granulocytes in RPD, we compared the proportion of granulocytes between RPD patients and controls. We found that the proportion of granulocytes was 2-fold higher in RPD patients compared to controls (Figure 5D). These findings suggest that granulocytes may play a role in the development of RPD, and CPD may be involved in this process.

The molecular mechanism of CPD gene in the pathogenesis of RPD in single-cell resolution

As the CPD gene is most relevant to the pathogenesis of RPD according to the aforementioned results, more in-depth molecular mechanism investigations in the single-cell resolution were conducted focusing on this gene. The analysis focused on granulocytes, and a total of 3,090 cells belonging to granulocyte clusters were selected for further study. Eight clusters were identified, and the CPD gene was found to be enriched in cluster 0 (c0), which was annotated as CD14+neutrophil ($p = 2.91 \times 10^{-34}$; Figure S7B).

To further investigate the molecular mechanisms underlying CPD gene involvement in RPD pathogenesis, we performed a correlation analysis on 3,090 granulocyte cells, revealing 1,608 positively correlated and 503 negatively correlated genes with CPD expression ($p < 0.05$). GO enrichment analysis revealed that the positively correlated genes were mainly associated with skeletal development processes, including bone resorption, osteoclast differentiation, osteoblast differentiation, and ossification (Figure 5E). In contrast, negatively correlated genes were mainly involved in immune response and fibroblast proliferation (Figure S8A), indicating that CPD may regulate the immune response and skeletal development processes during RPD pathogenesis. Furthermore, KEGG pathway analysis revealed significant enrichments in the NF-κB signaling pathway and the MAPK signaling pathway (Figure 5F), both of which have been implicated in the pathogenesis of RPD.[13,19] Although the role of MAPK signaling pathway in RPD pathogenesis remains unknown, its regulation of the inflammatory microenvironment may impact chondrocyte metabolism and contribute to joint degeneration.[20] Therefore, our findings suggest that CPD gene involvement in RPD pathogenesis may occur through these signaling pathways.

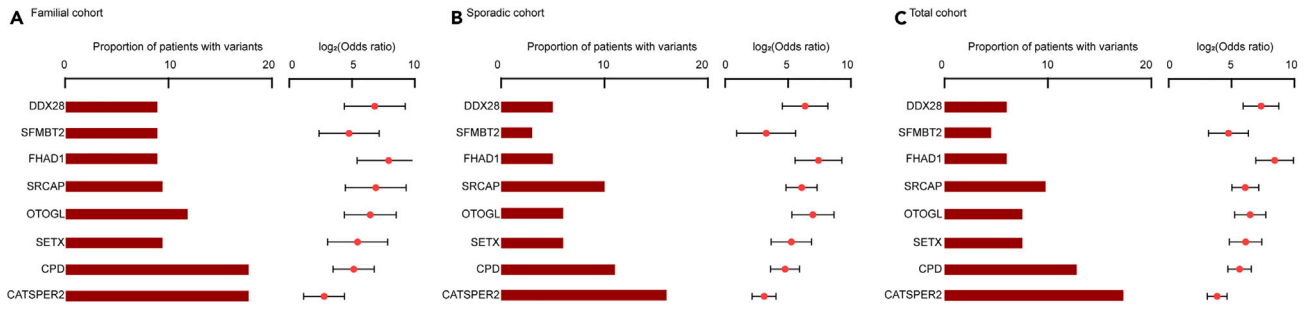


Figure 3. Frequency of deleterious variants in disease cohort versus controls

(A) The proportion of patients that carried the deleterious variants in the familial cohort (n = 33).

(B) The proportion of patients that carried the deleterious variants in the sporadic cohort (n = 100).

(C) The proportion of patients that carried the deleterious variants in the total cohort (n = 133). All the candidate genes were significantly enrichment in the disease cohort. The Odds ratio of the burden test was presented in the forest plot. The value of the Odds ratio was log2 scale to normalize.

DISCUSSION

Here, we present, to the best of our knowledge, the first WES study of a cohort of individuals with RPD. We identified eight novel genes that are associated with RPD and observed a significantly higher mutation rate and mutation burden of these genes in the RPD cohort compared to the controls. Moreover, patients who carried these RPD-associated genes exhibited more severe skeletal dysplasia phenotypes. In addition, we conducted scRNA-seq analysis of bone marrow samples from individuals with RPD, which revealed that CPD, one of the RPD-associated genes, was significantly enriched in the granulocytes. Our findings suggest that CPD influences the pathogenesis of RPD by regulating the immune response and skeletal development. Notably, the scRNA-seq results provide insights into the potential biological processes and signaling pathways that the CPD gene may be involved in.

The identification of candidate genes for RPD has been a long-standing challenge despite previous reports of the disease's heritability. Miller et al.[21] reported the first case of 12 individuals with RPD in 3 generations of a family with an autosomal dominant pattern in 1978. Although there have been sporadic studies reporting the genetic susceptibility of RPD,[10,11,22] no study has conducted WES in an RPD disease

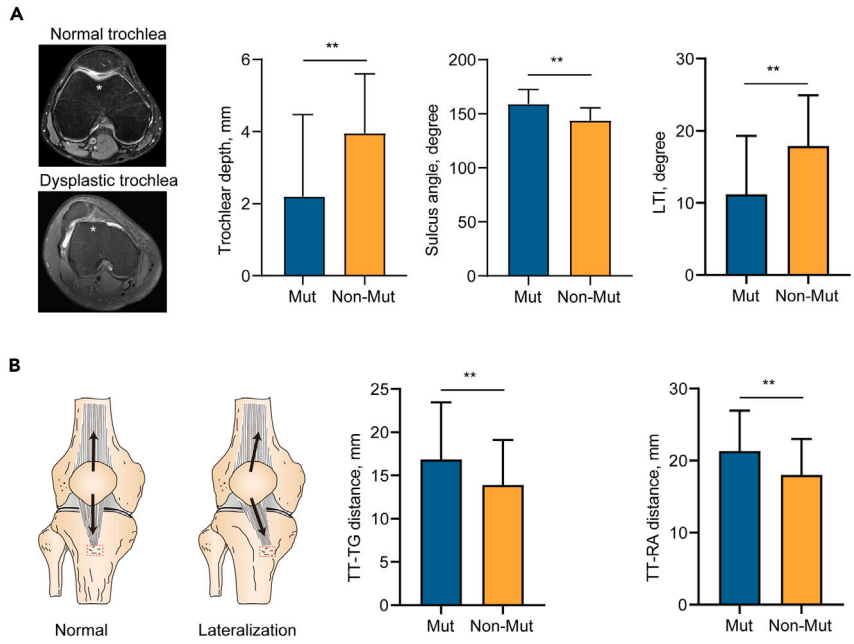


Figure 4. Skeletal deformities comparison between Mut group and non-Mut group

(A) The femoral trochlea development comparison. Normal trochlea shows a deeper trochlear groove; however, the dysplasia trochlea is flatter than normal (white star). The Mut group shows a shallower trochlear depth (TD, $p < 0.001$), greater sulcus angle (SA, $p < 0.001$), and smaller lateral trochlear inclination (LTI, $p < 0.001$), indicating more severe trochlea dysplasia.

(B) The tibial tubercle malposition evaluation. The Mut group shows greater tibial tubercle-trochlear groove (TT-TG) distance ($p = 0.007$) and tibial tubercle-Roman arch (TT-RA) distance ($p < 0.001$), indicating abnormal tibia development. ** indicates $p < 0.001$.

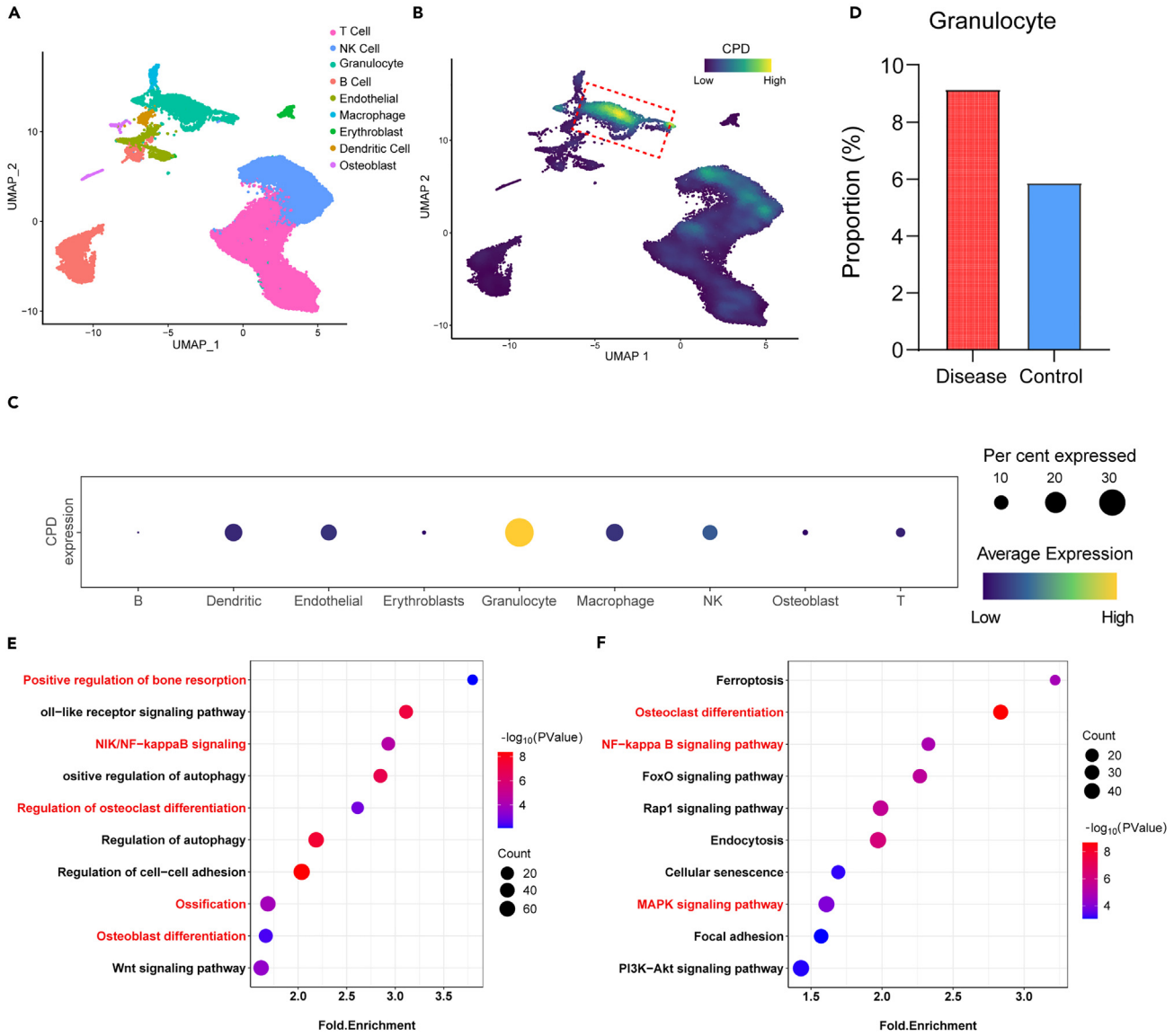


Figure 5. The single-cell landscape of bone marrow in RPD patients and the role of CPD gene in the pathogenesis

(A) Uniform manifold approximation and projection (UMAP) visualization of the clusters identified in the human bone marrow.

(B) Projection of CPD gene expression. Granulocyte clusters are outlined with a red pane.

(C) Percentage of CPD-expression cells and CPD expression levels.

(D) The difference in the granulocyte proportion between RPD patients and controls.

(E) The GO analysis of genes positively associated with the CPD expression.

(F) The KEGG pathway analysis of genes positively associated with the CPD expression.

cohort. Consequently, no convincing susceptibility genes or loci associated with RPD have been identified to date. To identify a convincing RPD-associated gene, we conducted a two-stage study. We first identified potentially associated variants in 33 cases with familial RPD and subsequently validated these variants in 100 sporadic cases. Our analysis identified eight variants with high confidence. The severe trochlear dysplasia phenotype in mutation gene carriers and the higher mutation burden of the corresponding genes in the RPD cohort than in controls supports the causality of the identified genes.

Our study also revealed the most relevant gene (CPD) with the RPD pathophysiology. Although RPD is known to affect bone, ligaments, and muscle,[23–25] it remains unclear which of these tissues is most closely related to the disease. To address this question, we assessed the expression of the identified genes in the three most relevant tissues (bone marrow, skeletal muscle, and cartilage) to determine the most relevant gene to RPD pathophysiology. Our analysis revealed that the CPD gene is the most relevant gene, with high expression in all three tissues. CPD belongs to the metallocarboxypeptidase family, and dysregulation of this family has been associated with various skeletal

phenotypes. For instance, inhibition of carboxypeptidase A can induce chondrocyte differentiation,[26] while carboxypeptidase E (CPE) is a modulator of osteoclasts, and CPE knockout mice exhibit decreased bone mineral density.[27,28] However, the role of CPD in skeletal disorders has yet to be investigated.

To the best of our knowledge, the use of scRNA-seq has not been previously reported in investigating the pathogenesis of RPD. Compared to traditional bulk RNA-seq, scRNA-seq offers higher sensitivity and specificity in identifying molecular alterations and allows for the evaluation of molecular mechanisms at a single-cell resolution.[29–31] Bone marrow provides a spatial environment for osteolineage cells, stromal cells, and immune cells,[32–34] and disruption of bone marrow physiology can impact osteogenic differentiation and mineralization.[35,36] Furthermore, bone marrow lesions are associated with abnormal hyperosteogeny, leading to OA.[37,38] Therefore, the bone marrow may play a significant role in the pathophysiology of RPD. Interestingly, our results suggest that the granulocyte may be a critical cell type involved in RPD pathogenesis through the CPD gene, which is highly expressed in bone marrow.

In conclusion, this study highlights the significant role of genetic factors in the development of RPD and provides novel insights into its etiology. The identification of the RPD-associated genes, especially the eight novel genes identified in this study, may provide crucial information for disease mechanisms understanding, genetic counseling, and precision medicine.

Limitations of the study

There are several limitations to this study. Firstly, the sample size of patients who underwent WES is relatively small, and all the patients were of Chinese population. Although the use of familial cases can identify disease-associated genes with higher effect sizes, further studies with larger sample sizes and diverse populations are still needed to identify more RPD-associated genes and validate the findings of this study. Moreover, animal experiments such as gene knockout models can provide more robust evidence of the association between the identified genes and RPD pathogenesis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study design and participants
- METHODS DETAILS
  - Joints hypermobility examination
  - Radiological examinations
  - Whole-exome sequencing and analysis
  - Functional enrichment analysis
  - Expression analysis
  - ScRNA-seq of human bone marrow
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109697.

ACKNOWLEDGMENTS

This work was supported by Ministerial Commission of Science and Technology: Special Project of Patellar Dislocation (JK-2022-07), National Natural Science Foundation of China (51920105006, 82002298), and Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C263). Part of the analysis was performed on the Computing Platform of the Center for Life Sciences of Peking University and supported by the High-Performance Computing Platform of Peking University.

AUTHOR CONTRIBUTIONS

Conceptualization, Z.X., J.Y., R.X., and J.Y.; methodology, Z.X., S.H., Y.F., R.X., H.W., and J.Y.; software programming, Z.X., S.H., O.L., and R.X.; validation, Z.X., S.H., H.W., R.X., and J.Y.; formal analysis, Z.X., S.W., Y.S., H.W., R.X., and J.Y.; investigation, Z.X., S.W., Y.S., C.X., H.Y., O.L., B.L., F.Y., B.X., H.W., R.X., and J.Y.; resources, C.X., H.Y., O.L., B.L., F.Y., B.X., H.W., R.X., and J.Y.; data curation, Z.X., S.W., H.W., R.X., and J.Y.; writing – original draft, Z.X. and S.H.; writing – review & editing, Z.X., S.W., Y.S., H.W., R.X., and J.Y.; visualization, Z.X., S.W., Y.S., C.X., H.Y., R.X., and J.Y.; supervision, H.W., R.X., and J.Y.; project administration, Z.X., S.H., Y.F., R.X., H.W., and J.Y.; funding acquisition, C.X., B.X., H.W., R.X., and J.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 18, 2023
Revised: February 5, 2024
Accepted: April 5, 2024
Published: April 8, 2024

REFERENCES

1. Lyons, J.G., Hudson, T.L., and Krishnamurthy, A.B. (2024). Epidemiology of patellar dislocations in the United States from 2001 to 2020: results of a national emergency department database. Physician Sportsmed. *52*, 26–35. https://doi.org/10.1080/00913847.2022.2156765.
2. Gravesen, K.S., Kallemose, T., Blønd, L., Troelsen, A., and Barfod, K.W. (2018). High incidence of acute and recurrent patellar dislocations: a retrospective nationwide epidemiological study involving 24.154 primary dislocations. Knee Surg. Sports Traumatol. Arthrosc. *26*, 1204–1209. https://doi.org/10.1007/s00167-017-4594-7.
3. Bhashyam, A.R., and Weaver, M.J. (2018). Knee pain after a fall. Br. Med. J. *360*, k775. https://doi.org/10.1136/bmj.k775.
4. Grant, C., Fick, C.N., Welsh, J., McConnell, J., and Sheehan, F.T. (2021). A Word of Caution for Future Studies in Patellofemoral Pain: A Systematic Review With Meta-analysis. Am. J. Sports Med. *49*, 538–551. https://doi.org/10.1177/0363546520926448.
5. Whittaker, J.L., Losciale, J.M., Juhl, C.B., Thorlund, J.B., Lundberg, M., Truong, L.K., Miciak, M., van Meer, B.L., Culvenor, A.G., Crossley, K.M., et al. (2022). Risk factors for knee osteoarthritis after traumatic knee injury: a systematic review and meta-analysis of randomised controlled trials and cohort studies for the OPTIKNEE Consensus. Br. J. Sports Med. *56*, 1406–1421. https://doi.org/10.1136/bjsports-2022-105496.
6. Fithian, D.C., Paxton, E.W., Stone, M.L., Silva, P., Davis, D.K., Elias, D.A., and White, L.M. (2004). Epidemiology and natural history of acute patellar dislocation. Am. J. Sports Med. *32*, 1114–1121. https://doi.org/10.1177/0363546503260788.
7. Danielsen, O., Poulsen, T.A., Eysturoy, N.H., Mortensen, E.S., Hölmich, P., and Barfod, K.W. (2023). Familial association and epidemiological factors as risk factors for developing first time and recurrent patella dislocation: a systematic review and best knowledge synthesis of present literature. Knee Surg. Sports Traumatol. Arthrosc. *31*, 3701–3733. https://doi.org/10.1007/s00167-022-07265-z.
8. Borochowitz, Z., Soudry, M., and Mendes, D.G. (1988). Familial recurrent dislocation of patella with autosomal dominant mode of inheritance. Clin. Genet. *33*, 1–4. https://doi.org/10.1111/j.1399-0004.1988.tb04257.x.
9. Alfaiz, A.A., Micale, L., Mandriani, B., Augello, B., Pellico, M.T., Chrast, J., Xenarios, I., Zelante, L., Merla, G., and Reymond, A. (2014). TBC1D7 mutations are associated with intellectual disability, macrocrania, patellar dislocation, and celiac disease. Hum. Mutat. *35*, 447–451. https://doi.org/10.1002/humu.22529.
10. Zhang, Q.H., Zhang, Y., He, R.X., Guo, H.M., and Wang, X.G. (2022). Anatomical characteristics and potential gene mutation sites of a familial recurrent patellar dislocation. BMC Med. Genom. *15*, 176. https://doi.org/10.1186/s12920-022-01330-9.
11. Kim, S.J., Yoon, J.S., and Hwang, I.T. (2022). A Novel Heterozygous ACAN Variant in a Short Patient Born Small for Gestational Age with Recurrent Patellar Dislocation: A Case Report. J. Clin. Res. Pediatr. Endocrinol. *14*, 481–484. https://doi.org/10.4274/jcrpe.galenos.2021.2021.0081.
12. Xu, C., Dong, Z., Ji, G., Yan, L., Wang, X., Li, K., Liu, J., Zhao, J., and Wang, F. (2022). RNA-seq based integrative analysis of potential crucial genes and pathways associated with patellar instability. Bioengineered *13*, 11402–11416. https://doi.org/10.1080/21655979.2022.2062528.
13. Lin, W., Kang, H., Dai, Y., Niu, Y., Yang, G., Niu, J., Li, M., and Wang, F. (2021). Early patellofemoral articular cartilage degeneration in a rat model of patellar instability is associated with activation of the NF-κB signaling pathway. BMC Muscoskel. Disord. *22*, 90. https://doi.org/10.1186/s12891-021-03965-8.
14. Malek, S., Reinhold, E.J., and Pearce, G.S. (2021). The Beighton Score as a measure of generalised joint hypermobility. Rheumatol. Int. *41*, 1707–1716. https://doi.org/10.1007/s00296-021-04832-4.
15. Steinberg, J., Southam, L., Fontalis, A., Clark, M.J., Jayasuriya, R.L., Swift, D., Shah, K.M., Brooks, R.A., McCaskie, A.W., Wilkinson, J.M., and Zeggini, E. (2021). Linking chondrocyte and synovial transcriptional profile to clinical phenotype in osteoarthritis. Ann. Rheum. Dis. *80*, 1070–1074. https://doi.org/10.1136/annrheumdis-2020-219760.
16. Xu, W.D., Huang, Q., and Huang, A.F. (2021). Emerging role of galectin family in inflammatory autoimmune diseases. Autoimmun. Rev. *20*, 102847. https://doi.org/10.1016/j.autrev.2021.102847.
17. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
18. Cao, Y., Li, L., Xu, M., Feng, Z., Sun, X., Lu, J., Xu, Y., Du, P., Wang, T., Hu, R., et al. (2020). The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals. Cell Res. *30*, 717–731. https://doi.org/10.1038/s41422-020-0322-9.
19. Xu, C., Ji, G., Chen, X., Yan, L., Liang, T., Liu, J., and Wang, F. (2022). Sclerostin antibody promotes bone formation through the Wnt/β-catenin signaling pathway in femoral trochlear after patellar instability. Connect. Tissue Res. 1–13. https://doi.org/10.1080/03008207.2022.2135507.
20. Zhang, Z., Yuan, L., Liu, Y., Wang, R., Zhang, Y., Yang, Y., Wei, H., and Ma, J. (2023). Integrated Cascade Nanozyme Remodels Chondrocyte Inflammatory Microenvironment in Temporomandibular Joint Osteoarthritis via Inhibiting ROS-NF-κB and MAPK Pathways. Adv. Healthcare Mater. *12*, e2203195. https://doi.org/10.1002/adhm.202203195.
21. Miller, G.F. (1978). Familial recurrent dislocation of the patella. J. Bone Joint Surg. Br. *60-b*, 203–204. https://doi.org/10.1302/0301-620x.60b2.659465.
22. Chan, C.J., Chau, Y.J., Woo, S.B., Luk, H.M., and Lo, I.F. (2018). Familial patellar dislocation associated with t(15;20)(q24;q13.1). J. Orthop. Surg. *26*, 2309499018777026. https://doi.org/10.1177/2309499018777026.
23. DeVries, C.A., Bomar, J.D., and Pennock, A.T. (2021). Prevalence of Trochlear Dysplasia and Associations with Patellofemoral Pain and Instability in a Skeletally Mature Population. J. Bone Joint Surg. Am. *103*, 2126–2132. https://doi.org/10.2106/jbjs.20.01624.
24. Xu, Z., Song, Y., Deng, R., Zhang, Z., Wang, H., and Yu, J.K. (2022). Pathological Thresholds of Segmental Femoral Torsion in Patients With Patellar Dislocation: Influence on Patellofemoral Malalignment. Orthop. J. Sports Med. *10*, 23259671221125218. https://doi.org/10.1177/23259671221125218.
25. Arrebola, L.S., Smith, T., Silva, F.F., de Oliveira, V.G.C., de Oliveira, P.R., Wun, P.Y.L., and Pinfildi, C.E. (2021). Hip and Knee Weakness and Ankle Dorsiflexion Restriction in Individuals Following Lateral Patellar Dislocation: A Case-Control Study. Clin. J. Sport Med. *31*, e385–e391. https://doi.org/10.1097/jsm.0000000000000815.
26. Kadouchi, I., Sakamoto, K., Tangjiao, L., Murakami, T., Kobayashi, E., Hoshino, Y., and Yamaguchi, A. (2009). Latexin is involved in bone morphogenetic protein-2-induced chondrocyte differentiation. Biochem. Biophys. Res. Commun. *378*, 600–604. https://doi.org/10.1016/j.bbrc.2008.11.111.
27. Cawley, N.X., Yanik, T., Woronowicz, A., Chang, W., Marini, J.C., and Loh, Y.P. (2010). Obese carboxypeptidase E knockout mice exhibit multiple defects in peptide hormone processing contributing to low bone mineral density. Am. J. Physiol. Endocrinol. Metab. *299*, E189–E197. https://doi.org/10.1152/ajpendo.00516.2009.
28. Kim, H.J., Hong, J., Yoon, H.J., Yoon, Y.R., and Kim, S.Y. (2014). Carboxypeptidase E is a novel modulator of RANKL-induced osteoclast differentiation. Mol. Cell. *37*, 685–690. https://doi.org/10.14348/molcells.2014.0179.
29. Su, M., Pan, T., Chen, Q.Z., Zhou, W.W., Gong, Y., Xu, G., Yan, H.Y., Li, S., Shi, Q.Z., Zhang, Y., et al. (2022). Data analysis guidelines for single-cell RNA-seq in

biomedical studies and clinical applications. Mil. Med. Res. *9*, 68. https://doi.org/10.1186/s40779-022-00434-8.

30. Pollen, A.A., Kilik, U., Lowe, C.B., and Camp, J.G. (2023). Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution. Nat. Rev. Genet. *24*, 687–711. https://doi.org/10.1038/s41576-022-00568-4.
31. Jia, Q., Chu, H., Jin, Z., Long, H., and Zhu, B. (2022). High-throughput single-cell sequencing in cancer research. Signal Transduct. Targeted Ther. *7*, 145. https://doi.org/10.1038/s41392-022-00990-4.
32. Hofbauer, L.C., Bozec, A., Rauner, M., Jakob, F., Perner, S., and Pantel, K. (2021). Novel approaches to target the microenvironment of bone metastasis. Nat. Rev. Clin. Oncol. *18*, 488–505. https://doi.org/10.1038/s41571-021-00499-9.
33. Matsushita, Y., Liu, J., Chu, A.K.Y., Tsutsumi-Arai, C., Nagata, M., Arai, Y., Ono, W., Yamamoto, K., Saunders, T.L., Welch, J.D., and Ono, N. (2023). Bone marrow endosteal stem cells dictate active osteogenesis and aggressive tumorigenesis. Nat. Commun. *14*, 2383. https://doi.org/10.1038/s41467-023-38034-2.
34. Wang, J., Zhang, Y., Cao, J., Wang, Y., Anwar, N., Zhang, Z., Zhang, D., Ma, Y., Xiao, Y., Xiao, L., and Wang, X. (2023). The role of autophagy in bone metabolism and clinical significance. Autophagy *19*, 2409–2427. https://doi.org/10.1080/15548627.2023.2186112.
35. Zhang, G., and Liu, J. (2021). Targeting senescent immune cells to rejuvenate the aging skeleton. Cell Metabol. *33*, 1903–1905. https://doi.org/10.1016/j.cmet.2021.09.005.
36. van Gastel, N., and Carmeliet, G. (2021). Metabolic regulation of skeletal cell fate and function in physiology and disease. Nat. Metab. *3*, 11–20. https://doi.org/10.1038/s42255-020-00321-3.
37. Walsh, D.A., Sofat, N., Guermazi, A., and Hunter, D.J. (2023). Osteoarthritis Bone Marrow Lesions. Osteoarthritis Cartilage *31*, 11–17. https://doi.org/10.1016/j.joca.2022.09.007.
38. Hansen, R.T., Chenu, C., Sofat, N., and Pitsillides, A.A. (2023). Bone marrow lesions: plugging the holes in our knowledge using animal models. Nat. Rev. Rheumatol. *19*, 429–445. https://doi.org/10.1038/s41584-023-00971-z.
39. Lin, W., Kang, H., Niu, Y., Niu, J., Fan, C., Feng, X., and Wang, F. (2022). Cartilage degeneration is associated with activation of the PI3K/AKT signaling pathway in a growing rat experimental model of developmental trochlear dysplasia. J. Adv. Res. *35*, 109–116. https://doi.org/10.1016/j.jare.2021.04.006.
40. Xu, Z., Zhao, P., Song, Y., Wang, H., Zhou, A., and Yu, J.K. (2022). Reliability of the Tibial Tubercle-Roman Arch Distance for Evaluating Tibial Tubercle Malposition and Predicting Patellar Dislocation via Magnetic Resonance Imaging. Orthop. J. Sports Med. *10*, 23259671221118561. https://doi.org/10.1177/23259671221118561.
41. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698.
42. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
43. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly *6*, 80–92. https://doi.org/10.4161/fly.19695.
44. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814. https://doi.org/10.1093/nar/gkg509.
45. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *76*. Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76.
46. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47*, D886–D894. https://doi.org/10.1093/nar/gky1016.
47. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.
48. Ruoss, S., Walker, J.T., Nasamran, C.A., Fisch, K.M., Paez, C.J., Parekh, J.N., Ball, S.T., Chen, J.L., Ahmed, S.S., and Ward, S.R. (2021). Strategies to Identify Mesenchymal Stromal Cells in Minimally Manipulated Human Bone Marrow Aspirate Concentrate Lack Consensus. Am. J. Sports Med. *49*, 1313–1322. https://doi.org/10.1177/0363546521993788.
49. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
50. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
51. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44. https://doi.org/10.1038/nbt.4314.
52. Alquicira-Hernandez, J., and Powell, J.E. (2021). Nebulosa recovers single cell gene expression signals by kernel density estimation. Bioinformatics *37*, 2485–2487. https://doi.org/10.1093/bioinformatics/btab003.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Peripheral blood of patients	Cohort sample in article	N/A
Bone marrow of patients	Cohort sample in article	N/A
Cartilage of mice	Cohort sample in article	N/A
Critical commercial assays		
SureSelect Human All Exon kit V6	Agilent Technologies, Santa Clara, CA, USA	https://www.agilent.com/
Deposited data		
Bone marrow single-cell RNA sequencing data	This paper	Geo: GSE244724
Bone marrow and skeletal muscle expression data	Human Protein Atlas database	https://www.proteinatlas.org/
gnomAD	Genome Aggregation Database	https://gnomad.broadinstitute.org/
ChinaMAP	mBiobank	http://www.mbiobank.com/
Software and algorithms		
R 4.1.0	R project	https://www.r-project.org/
BWA version 0.7.12	Heng Li	https://github.com/lh3/bwa
SnpEff software version 4.3.1	Github	https://pcingola.github.io/SnpEff/
CADD version 1.6	Combined Annotation Dependent Depletion	https://cadd.gs.washington.edu/
SIFT	Pauline C Ng	http://blocks.fhcrc.org/sift/SIFT.html
PolyPhen2	Ivan Adzhubei	http://genetics.bwh.harvard.edu/pph2/
ClusterProfiler package	Github	https://guangchuangyu.github.io/software/clusterProfiler/
Hisat2 version 2.2.1	Github	https://daehwankimlab.github.io/hisat2/
Stringtie version 2.2.1	Github	https://github.com/gpertea/stringtie
Cell Ranger 6.0.1	10x genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
Seurat v4.0.6	satijalab	https://github.com/satijalab/seurat
Harmony (v1.0)	Github	https://github.com/immunogenomics/harmony
Nebulosa R package (v1.0.0)	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/Nebulosa.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for raw data and codes should be directed to and fulfilled upon reasonable request by the Lead contact Jia-Kuo Yu (yujiakuo@126.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The generated RNA-sequencing data has been deposited in GEO database and are publicly available as of the date of publication. Accession identifications are listed in the key resources table. This study (NCT04997538) has been registered in https://clinicaltrials.gov/.

- This study did not generate any original code.
- The raw whole exome sequencing data and any additional information are available from the lead contact upon reasonable request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design and participants

This prospective cohort study enrolled 533 consecutive Han Chinese patients with a diagnosis of RPD between March 2019 and May 2022 at the Sports Medicine Department of Peking University Third Hospital. Clinical information, including onset age, family history, and physical examination, was recorded prospectively. The study population was divided into two groups: the first group comprised 400 RPD patients enrolled between March 2019 and March 2021, and the second group comprised 133 patients enrolled between April 2021 and May 2022. The first group only collected clinical information to evaluate the potential hereditable tendency of the disease, while the second group provided both clinical information and peripheral blood samples for genetic analysis. Single-cell sequencing was performed on bone marrow from RPD patients to explore the molecular mechanism of this disease in the single-cell resolution. This has been registered on ClinicalTrial.gov (ID: NCT04997538). The study design was presented in Figure 1.

METHODS DETAILS

Joints hypermobility examination

Joint hypermobility was evaluated using the Beighton score,[14] which involves five tests: first-finger opposition, fifth-finger extension, elbow extension, knee extension, and forward bending. Patients with Beighton scores of 4 or greater were identified as having joint hypermobility.

Radiological examinations

The imaging modalities used in this study included X-ray, computerized tomography (CT), and magnetic resonance imaging (MRI), which were performed routinely. The data of examination was stored in the Electronic Medical Systems. Trochlear dysplasia was evaluated by measuring Dejour classification, trochlear depth, sulcus angle, and lateral trochlear inclination (LTI).[39] The tibial tubercle lateralization was evaluated by measuring the tibial tubercle-trochlear groove (TT-TG) distance and the tibial tubercle-Roman arch (TT-RA) distance.[40]

Whole-exome sequencing and analysis

Whole-exome sequencing was performed on a total of 133 patients, including 33 familial RPD patients in the discovery cohort and 100 sporadic RPD patients in the validation cohort. Genomic DNA was extracted from peripheral blood samples using standard protocols, and after quality control, the DNA was fragmented by Covaris and captured using the SureSelect Human All Exon kit V6 (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed on the Illumina Novaseq 6000 platform. The sequence reads were mapped to the human reference genome (GRCh38) using Burrows-Wheeler Aligner (BWA version 0.7.12),[41] and the variants were called using GATK pipline.[42] Annotation of the variants was performed using SnpEff software (version 4.3.1).[43] The population allele frequency data were derived from gnomAD.[17] The deleteriousness of the variants on protein structure was predicted using SIFT,[44] PolyPhen2,[45] and Combined Annotation-Dependent Depletion score (CADD version 1.6).[46]

We identify the candidate genes with the variants inherited dominantly. Based on the prevalence of RPD, we used the following criteria to identify candidate variants. (1) variants with allele frequency less than 0.01 in gnomAD and (2) nonsynonymous variants with CADD Phred score $\geq$15.

Functional enrichment analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the clusterProfiler package.[47] For the WES data, variants with a frequency greater than 5% in RPD patients (at least 7 carriers) and less than 1% in the gnomAD database were included in the analysis. Enrichment analysis results with adjusted p values less than 0.05 were considered statistically significant.

Expression analysis

The expression of the candidate genes in skeletal muscle and bone marrow was obtained from the Human Protein Atlas database (https://www.proteinatlas.org/), while the expression of genes in cartilage was obtained from in-house bulk RNA-seq data. Cartilage samples were harvested from five four-week-old male C57BL/6 mice, and total RNA was extracted using the Qiagen RNeasy mini kit (Qiagen, Germany). The cDNA libraries were sequenced on a BGIseq-500 sequencer (BGI-Shenzhen, China), and the raw data were processed by removing reads containing adapters, reads containing poly-N, and low-quality reads. The clean reads were aligned to the reference genome using Hisat2 (version 2.2.1). Finally, the expression of genes was quantified using Stringtie software (version 2.2.1) and normalized to transcript per million (TPM).

ScRNA-seq of human bone marrow

Bone marrow samples were aspirated from the distal femur of three RPD patients recruited at Peking University Third Hospital. The bone marrow scRNA-seq data of three non-RPD controls were obtained from the previously published study.[48] The bone marrow samples were collected into heparin tubes and density gradient centrifugation was used to isolate bone marrow cells. The single-cell RNA libraries were prepared using Chromium Single Cell 5′ Library and Gel Bead Kit. The sequence reads were mapped to the GRCH38 human reference genome using Cell Ranger 6.0.1 (10x Genomics).

After quality control, we performed downstream data analysis using the R package Seurat (v4.0.6) for dimensionality reduction, clustering, differential expression analysis, and visualization.[49] We log-normalized the data with a scale factor of 10,000 and identified highly variable features using the FindVariableFeature function with nfeatures = 2,000. We then used principal component analysis on the scaled data and integrated data from different patients using the R package Harmony (v1.0).[50] For dimensionality reduction, we used the uniform manifold approximation and projection (UMAP)[51] and annotated cell clusters based on the expression of canonical marker genes (e.g., CD3D for T cells, GNLY for NK cells, CD79A for B cells, etc.). To clarify the cell-type-specific expression of RPD-associated genes, we generated gene-weighted density plots using the Nebulosa R package (v1.0.0).[52] Additionally, we used Pearson correlation coefficient to measure the expression correlation between genes, and the functional enrichment analysis was performed using genes that showed a significant correlation with the identified genes ($p < 0.05$).

QUANTIFICATION AND STATISTICAL ANALYSIS

Fisher's exact test was conducted to compare the frequency of variants between RPD patients and the East Asia individuals without RPD in the gnomAD[17] database, and Chinese individuals from the ChinaMAP[18] database. Student's t test or Mann-Whitney U test is used to compare the quantitative data, such as age and tibial tubercle lateralization, between different groups. The chi-square test is used to compare the qualitative data, such as gender and the proportion of trochlear dysplasia, between groups. Odds ratios (ORs) and 95% confidence intervals (CIs) are calculated and a two-tailed p value less than 0.05 is considered statistically significant.