




# Genomic alterations of dermatofibrosarcoma protuberans revealed by whole-genome sequencing

Cong Peng <sup>1,2,3,4</sup> Xingxing Jian,<sup>1,4,5</sup> Yang Xie,<sup>6</sup> Lingfeng Li,<sup>1,2,3,4</sup> Jian Ouyang,<sup>5</sup> Ling Tang,<sup>1,2,3,4</sup> Xu Zhang,<sup>1,2,3,4</sup> Juan Su <sup>1,2,3,4</sup> Shuang Zhao,<sup>1,2,3,4</sup> Hong Liu,<sup>1,2,3,4</sup> Mingzhu Yin,<sup>1,2,3,4</sup> Dan Wu,<sup>5</sup> Miaojian Wan <sup>6</sup>, Lu Xie<sup>5</sup> and Xiang Chen<sup>1,2,3,4</sup>

<sup>1</sup>Department of Dermatology, Xiangya Hospital, Central South University, Changsha, Hunan, China

<sup>2</sup>Hunan Key Laboratory of Skin Cancer and Psoriasis, Changsha, Hunan, China

<sup>3</sup>Hunan Engineering Research Center of Skin Health and Disease, Changsha, Hunan, China

<sup>4</sup>National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Changsha, Hunan, China

<sup>5</sup>Shanghai-MOST Key Laboratory of Health and Disease Genomics, Institute for Genome and Bioinformatics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China

<sup>6</sup>Department of Dermatology, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

## Summary

### Correspondence

Xiang Chen.

Email: chenxiangck@126.com

### Accepted for publication

30 December 2021

### Funding sources

This work was supported by the National Natural Science Foundation of China (81620108024, 82073458, 81974476, 81830096), the Science and Technology Innovation Program of Hunan Province (2021RC4013), the Shanghai Municipal Health Commission Collaborative Innovation Cluster Project (No. 2019CXJQ02) and the Overseas Expertise Introduction Project for Discipline Innovation (111 Project, No. B20017).

### Conflicts of interest

The authors declare they have no conflicts of interest. C.P., M.W., L.X. and X.C. made an equal contribution. C.P., X.J., Y.X. and L.L. made an equal contribution.

### Data availability

The raw sequence data for all samples used in this study have been deposited in the public database of the National Omics Data Encyclopedia under project number OEP001296. These publicly deposited data (with controlled access on reasonable request) are available at <https://www.biosino.org/node/review/detail/OEV000137?code=CGCLNOL2>.

DOI 10.1111/bjd.20976

**Background** Dermatofibrosarcoma protuberans (DFSP) is a rare and marginal cutaneous sarcoma of intermediate-grade malignancy, for which the genomic landscape remains unclear. Understanding the landscape of DFSP will help to further classify the genomic pathway of malignant development in soft tissue.

**Objectives** To identify the comprehensive molecular pathogenesis of DFSP.

**Methods** In this study, the comprehensive genomic features, with 53 tumour-normal pairs of DFSP, were revealed by whole-genome sequencing.

**Results** The mutational signature 1 (C > T mutation at CpG dinucleotides) is featured in DFSP, resulting in higher mutations in DNA replication. Interestingly, the recurrence of DFSP is correlated with low tumour mutation burden. Novel mutation genes in DFSP were identified, including *MUC4/6*, *KMT2C* and *BRCA1*, and subsequently, three molecular subtypes of DFSP were classified on the basis of *MUC4* and *MUC6* mutations. Various structural aberrations including genomic rearrangements were identified in DFSPs, particularly in 17q and 22q, which cause oncogene amplification (*AKT1*, *SPHK1*, *COL1A1*, *PDGFβ*) or tumour suppressor deletion (*CDKN2A/B*). In addition to gene fusion of *COL1A1-PDGFβ* [t(17;22)], we identified gene fusion of *SLC2A5-BTBD7* [t(1;14)] in DFSP through whole-genome sequencing, and verified it experimentally. Enrichment analysis of altered molecules revealed that DNA repair, cell cycle, phosphoinositide 3-kinase and Janus kinase pathways were primarily involved in DFSP.

**Conclusions** This is the first large-scale whole-genome sequencing for DFSP, and our findings describe the comprehensive genomic landscape, highlighting the molecular complexity and genomic aberrations of DFSP. Our findings also provide novel potential diagnostic and therapeutic targets for this disease.

### What is already known about this topic?

- Chromosomal translocation between chromosome 17 and chromosome 22 is the main feature in the pathogenesis of dermatofibrosarcoma protuberans (DFSP).

### What does this study add?

- We describe the comprehensive genomic landscape of DFSP, highlighting the molecular complexity and genomic aberrations.
- Our findings provide novel potential diagnostic and therapeutic targets for this disease.

**What is the translational message?**

- Our study revealed novel molecular subtypes of DFSP based on genetic mutations, which benefits precision diagnosis.
- We also found oncogene amplification, including *AKT1* and *SPHK1*, which provides novel potential target molecules for further DFSP treatment.
- In addition to gene fusion of *COL1A1-PDGFβ*, we identified a novel gene fusion of *SLC2A5-BTBD7* in DFSP, which is a novel potential diagnostic and therapeutic target for this disease.

Dermatofibrosarcoma protuberans (DFSP) is a type of cutaneous soft tissue sarcoma, which is a rare malignant tumour with an incidence of approximately four per million.<sup>1</sup> In studies of the pathogenesis of DFSP, the whole spectrum of genomic characteristics is lacking. To date, studies have revealed that chromosomal translocation between chromosome 17 and chromosome 22 is the main feature in the pathogenesis of DFSP.<sup>2–4</sup> The second exon of platelet-derived growth factor- $\beta$  (*PDGFβ*) on chromosome 22 was fused to the collagen 1  $\alpha$ 1 (*COL1A1*) gene on chromosome 17, while various breakpoints of *COL1A1* from exon 6 to 49 have been observed.<sup>2,5,6</sup> *PDGFβ* is documented with the cellular homologue of the *v-sis* oncogene, which is involved in pathogenesis of diverse tumours including sarcoma.<sup>7</sup> Evidence shows that repressor elements in *PDGFβ* exon 1 negatively regulate its expression. Therefore, t(17;22) results in replacement of *PDGFβ* exon 1 by *COL1A1* elements; thus, a *COL1A1-PDGFβ* fusion oncogene leads to distinctly upregulated *PDGFβ* transcription. In addition to *COL1A1-PDGFβ* fusion, other fusion events were observed in DFSP, such as t(2;17), t(17;22), t(9;22) and t(5;8).<sup>8–11</sup>

In this study, 53 biopsy specimens from patients with DFSP and their paired blood samples were used for whole-genome sequencing (WGS), in which we identified a number of novel mutant genes, such as *MUC4/6* and *KMT2C*, and several significant segments, such as 17q25, 22q13. We also identified various structural alterations of tumour-related genes such as *SPHK1* and *BRCA1*, and the fusion gene *SLC2A5-BTBD7*. Overall, our findings highlight novel mechanisms of DFSP pathogenesis, particularly structural alterations, which help to identify potential diagnostic and therapeutic molecules.

## Materials and methods

### Patient characteristics and sample collection

All samples were mainly collected from the Department of Dermatology, The Third Affiliated Hospital, Sun Yat-sen University, and the sampling procedure was approved by the Medical Ethics Committee of The Third Affiliated Hospital, Sun Yat-sen University. Written informed consent was obtained from each patient. Detailed information for the participants is presented in Table S1a (see Supporting Information).

Biopsies were performed on patients with clinical suspicion of DFSP, and the diagnoses were made by three different pathologists. Haematoxylin and eosin staining and CD34 immunohistochemistry (IHC) were performed for each tumour and all samples were positive for CD34 (Table S1b; see Supporting Information). Furthermore, IHC for additional markers such as S100, smooth muscle actin, desmin and vimentin was performed for differential diagnostic purposes (Table S1b). Patients were considered eligible if they had a diagnosis of primary DFSP and had not received chemotherapy or radiotherapy treatment before the operation.

### Whole-genome sequencing

All samples were processed for whole-genome paired-end sequencing using Illumina HiSeq X10 Instruments (Novogene Bioinformatics Technology Co., Ltd., Tianjin, China). The sequencing data were applied to analyse single-nucleotide variants (SNVs), copy-number variants, structural variants and fusion genes. Full details are provided in Appendix S1 (see Supporting Information).

### Validation of amplification genes by polymerase chain reaction

Copy numbers of candidate genes were quantified for further validation in an ABI 7500 real-time polymerase chain reaction (PCR) system (Applied Biosystems, Waltham, MA, USA). A total of 20  $\mu$ L reaction volume was used with 20 ng of DNA from each sample, 2\*SYBR Green Mix and primers (10  $\mu$ mol<sup>-1</sup>). Each DNA sample was analysed using quantitative PCR in triplicate. The relative copy number of candidate genes was calculated by dividing the number of candidate gene copies by the number of albumin gene copies. The following primers were used in this study: *SPHK1* (F: CAGGGAAT-GACGCCGGTG; R: TCCTTATCGGTGCTGCCAG) and *ITGB4* (F: GGGACGAGGATGACGACTGC; R: CACCAGACAGTGCTGT TGG).

### Immunohistochemistry

Tissue samples from patients who were diagnosed with DFSP were fixed with 4% paraformaldehyde. Slides with 4- $\mu$ m

sections were baked at 65°C for 2 h and then rehydrated with ethanol (gradient concentration); 3% hydrogen peroxide was used to inhibit the endogenous peroxidase for 10 min after antigen retrieval. Tissues were then incubated with primary antibody SPHK1 (1 : 200) (Cell Signaling Technology, Danvers, MA, USA) and P-stat3 (Cell Signaling Technology, 1 : 200) overnight at 4°C. Slides that were well incubated with primary antibody were taken out from the refrigerator and incubated with horseradish peroxidase-conjugated secondary antibody at room temperature for 1 h. Finally, slides stained with DAB were counterstained with haematoxylin and mounted with neutral balsam.

### Fluorescence *in situ* hybridization test of *COL1A1-PDGFβ* fusion gene in dermatofibrosarcoma protuberans tissues

All samples from patients were subjected to fluorescence *in situ* hybridization (FISH) analysis using the *COL1A1-PDGFβ* fusion gene t(17; 22) probe (Anbiping, Guangzhou, China). The fluorescent signal was observed using Olympus BX51 Microscopes (original magnification × 1000) (Olympus, Tokyo, Japan) and at least 200 cells were counted for each sample. Samples where at least 10% of tumour cells exhibited fusion signalling were considered to be positive samples.

### Validation of *SLC2A5-BTBD7* fusion gene by Sanger sequencing

DNA was extracted from the biopsy tissues of several patients (T27, T29, T32, T36, T28, T37) using QIAamp DNA MiNi Kit (Qiagen, Hilden, Germany). The target gene was fused by the *BTBD7* gene on chromosome 14 to *SLC2A5* on chromosome 1. The forward primer in exon 2 of *SLC2A5*, and the reverse primer in exon 7 of *BTBD7* were designed according to the sequence before and after the breakpoint identified by WGS (F: GGATTTCCCTATGTTGC; R: TGAAAGTCCACTCTTGAC). PCR amplification using KOD FX DNA polymerase (Toyobo Inc., Tokyo, Japan) consisted of an initial cycle at 94°C for 10 min followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s. The final extension was performed at 72°C for 10 min. The amplification products were analysed by electrophoresis on 1.0% agarose gels and sequenced by Sangon Biotech (Sangon, Shanghai, China). Subsequently, the sequences were applied to the Basic Local Alignment Search Tool for the corresponding gene on the National Center for Biotechnology Information database.

## Results

### Genomic somatic alterations in dermatofibrosarcoma protuberans

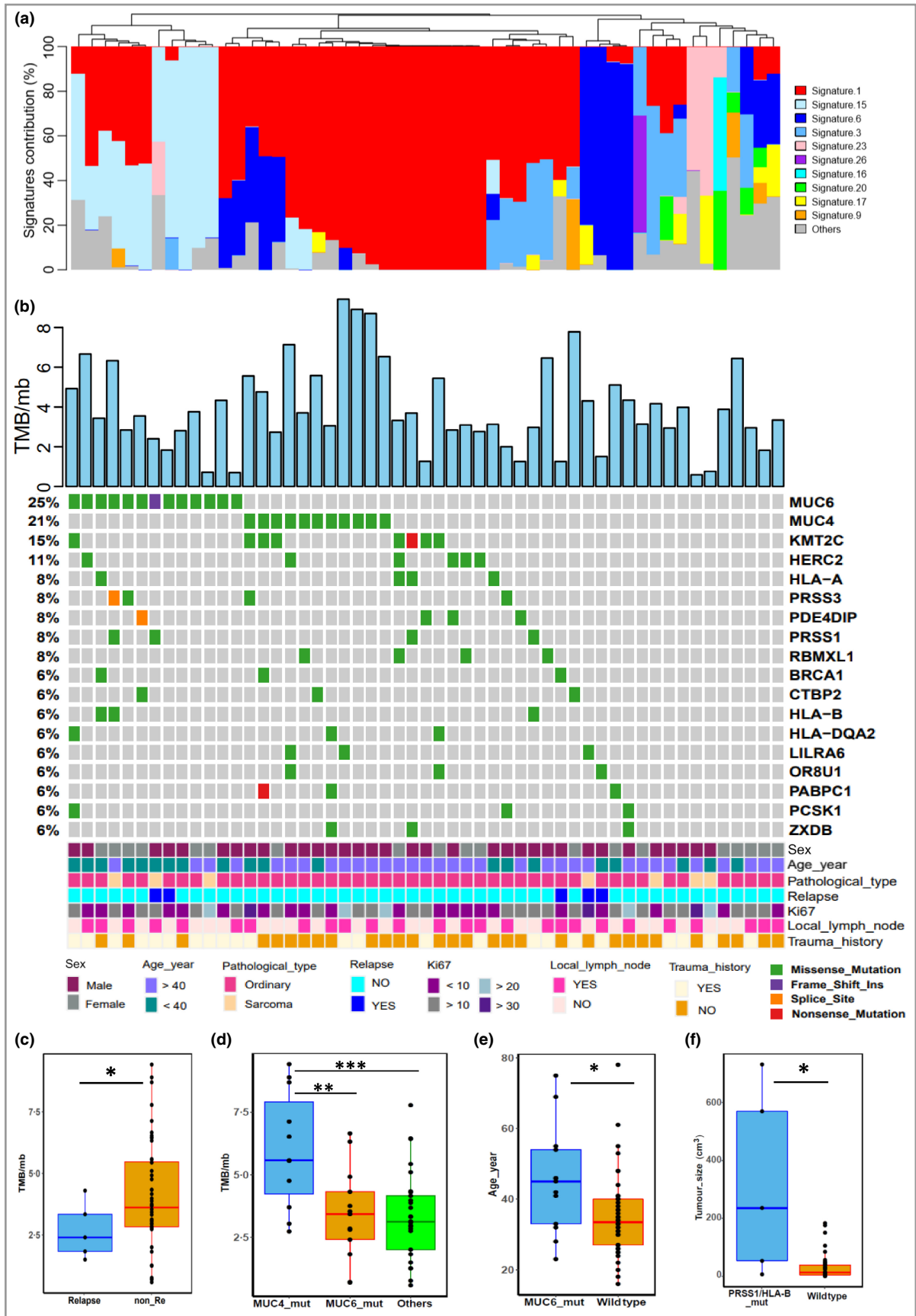
In this study, we performed WGS of the paired tumour-blood samples collected from 53 eligible patients, whose histopathology consisted of 46 ordinary samples and seven

fibrosarcomatous variant types. At subsequent follow-up, five of these patients had experienced relapse, two of whom had a fibrosarcomatous variant type (Table S1a).

Here, a total of 1255 somatic variants were detected, ranging from two to 84 variants per sample, with a median of 20 variations, including SNVs and small insertions and deletions (Figure S1d, Table S2a; see Supporting Information). SNVs accounted for 96% of somatic mutations, ranging from two to 78 variants per sample, with a median of 18. The six different SNV classes (T > G, T > C, T > A, C > T, C > G, C > A) were counted separately (Figure S1c), and the proportion of different SNV classes was calculated separately in all samples and in each sample, showing the two major SNV classes to be C > T and T > G, which accounted for 27% and 23% in all mutation positions, respectively (Figure S2a, c, Table S2b; see Supporting Information). As the nucleotide substitutions provide insight into the mutational processes that shaped the cancer genome, the mutational signatures were analysed further and signature 1 was shown to be a pivotal mutation type in DFSP, followed by signatures 3, 6 and 15 (Figure 1a, Table S2c). Signature 1, characterized by C > T mutation at NpCpG trinucleotides, is the result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine.

Compared with the 33 solid tumours in the The Cancer Genome Atlas database, the number of mutations per sample in DFSP was relatively low and similar to that found in the SARC tumour (Figure S3; see Supporting Information). Furthermore, the tumour mutation burden (TMB) of DFSP samples was separately evaluated, ranging from 0.58 to 9.42 mutations per megabase, with a median of 3.43 mutations per megabase sequenced [Figure 1b (top panel), Table S2d]. Interestingly, a significantly lower TMB was found in the samples from patients who had relapsed (Figure 1c), while no significant difference in TMB was found between fibrosarcomatous variants and ordinary samples, suggesting that a low TMB in DFSP may be associated with immune escape and not with pathological type.

In addition, 958 somatic mutant genes were detected in the DFSP cohort (Table S2e); 47 genes with mutations in at least three samples were detected (> 5%), which were subjected to further analysis. In particular, *MUC6* and *MUC4* showed high mutational frequency, accounting for 25% and 21% of patients, respectively. Subsequently, six genes were found to have been included in oncoKB,<sup>12</sup> an evidence-based oncology knowledge base pertaining to somatic mutations and structural alterations, such as *KMT2C* (15%), *PDE4DIP* (8%), *PRSS1* (8%), *HLA-A* (8%), *HLA-B* (6%) and *BRCA1* (6%). Through positional clustering of variants using the oncodriveCLUST algorithm,<sup>13</sup> 11 genes with mutational frequency greater than 5% were identified as driver genes (Table S2f). Taken together, as shown in Figure 1b (bottom panel), these findings lead us to speculate that these mutational genes may be related to DFSP tumorigenesis. Among these genes, we observed a significant mutual exclusion between *MUC4* and *MUC6*, and a significant cooccurrence between *PRSS1* and *HLA-B* (Figure S4; see





**Figure 1** Genomic somatic alterations in dermatofibrosarcoma protuberans (DFSPs). (a) Mutation signatures with the top 10 contributions. The point mutations identified in the DFSPs were analysed for mutational signatures using deconstructSigs. (b) Tumour mutation burden (TMB) per sample (top) and mutational waterfall of driver genes referred from oncoKB and identified by oncodriveCLUST (bottom). (c) Comparison of TMB for relapsed and nonrelapsed samples. (d) Comparison of TMB for MUC4 mutation, MUC6 mutation, and other samples. (e) Comparison of age between patients with MUC4 mutation and wildtype individuals. (f) Comparison of tumour size for patients with PRSS1/HLA-B-mutation and wildtype individuals. Significance was determined by using two-tailed Wilcoxon rank sum test (\* $P < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

Supporting Information). Interestingly, when MUC4 and MUC6 were taken as molecular characteristics of DFSP sample typing, our findings showed that mutation of MUC4 is associated with a high TMB compared with mutation of MUC6 and nonmutation of MUC4 and MUC6 (Figure 1d). Several recent studies showed that MUC4, MUC16 and the TTN gene mutations correlate with prognosis and predicted TMB and immunotherapy efficacy in gastric cancer and pan-cancer cohorts.<sup>14–16</sup> However, we did not detect MUC4, MUC16 and TTN mutation in the five patients who relapsed, suggesting that mutations in these genes may be used as biomarkers of relapse in DFSP. Furthermore, MUC6 mutation is closely related to the patients' age (Figure 1e). Given the significant cooccurrence of PRSS1 and HLA-B, we found that mutation of PRSS1 plus HLA-B is remarkably relevant to tumour size (Figure 1f). Additionally, genes with mutational frequency greater than 5% were taken to perform gene ontology (GO) enrichment on biological process (BP), molecular function (MF) and cell component (CC) (Figure S5a–b, Table S3a–c; See Supporting Information), showing that these genes (MUC4, MUC6 and MUC16) were mainly involved in immune response and protein glycosylation.

### Copy-number variation landscape and recurrently amplified genes in dermatofibrosarcoma protuberans

Genome-wide copy-number variations (CNVs) of sequence segments in each paired sample were separately evaluated using Sequenza (<https://cran.r-project.org>),<sup>17</sup> and 27 significant segments were identified in the cohort by GISTIC2.0 (<https://www.genepattern.org>),<sup>18</sup> including 16 amplification peaks (APs) and 11 deletion peaks (DPs) (Figure S6a, Table S4a; see Supporting Information). Subsequently, the genes mapping to the significant CNV segments in each tumour sample were summarized, with the number of CNV genes ranging from one to 2416, with a median of 1630 and a median of 1435 amplification genes and five deletion genes (Figure S6a, top panel).

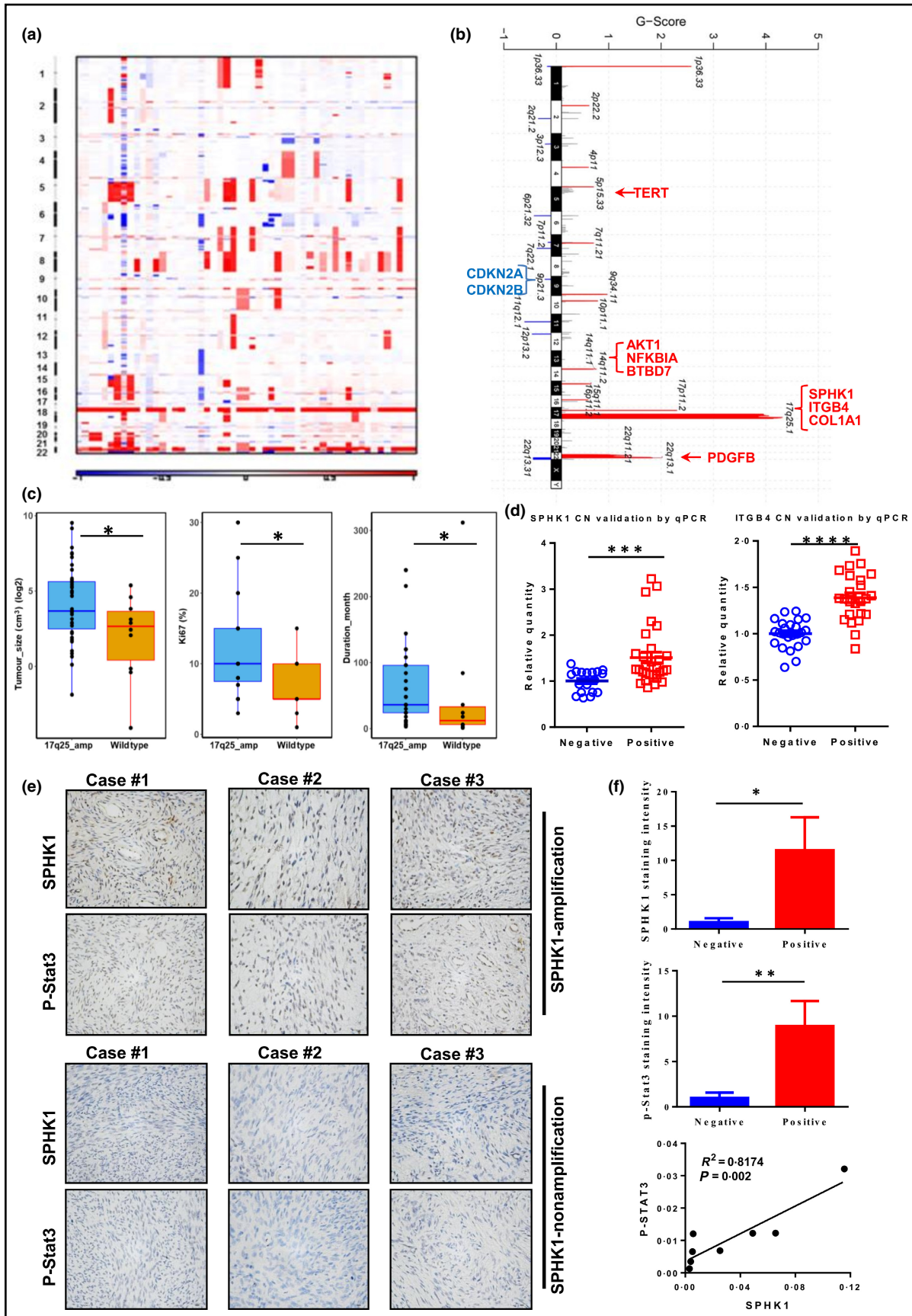
As shown in Figure 2a, the raw copy number revealed a large amount of amplified CNV in DFSP, especially in chromosomes 17 and 22. Meanwhile, the amplified segments of 17q25, 22q13 and 22q11 were detected in 81%, 79% and 75% of the samples in the cohort, respectively (Figure S6a, bottom panel). Significant genes located in APs and DPs were also identified (Table S4b, c). In particular, as shown in Figure 2b, we discovered that several identified amplified segments cover some well-known oncodriver genes, such as TERT (5p15.22), AKT1 (14q11.2), NFKBIA (14q11.2), BTBD7

(14q11.2), SPHK1 (17q25.1), ITGB4 (17q25.1) COL1A1 (17q25.1) and PDGF $\beta$  (22q13.1), and that the deletion segment covers tumour suppressor genes, including CDKN2A (9p21.3) and CDKN2B (9p21.3). These findings suggest that the CNVs of these fragments may be associated with DFSP tumorigenesis. We further found that the amplified 17q25 was positively related to tumour size, Ki67 expression and duration (Figure 2c), and that amplified 5p15, 4p11, 15q11 and 16p11 were also relevant to Ki67 expression in DFSP (Figure S6b). The identified amplified genes in DFSP were subjected to GO enrichment analysis, and the results showed that they are mainly involved in skin epidermal cell differentiation, mitotic cell cycle phase transition, and epidermal growth factor (Figure S7, Table S5a–c; see Supporting Information). However, no significant BP, MF or CC with Benjamini–Hochberg (BH)-adjusted P-value cutoff of 0.05 was found based on the genes mapped by significant deletion segments.

Moreover, copy-number alterations of SPHK1 and ITGB4 located in 17q25 were validated by quantitative PCR, which confirmed the amplification of 17q25 identified by WGS in DFSP (Figure 2d). Stat3 is a downstream molecule of the SPHK1 regulation signalling pathway, and p-Stat3 was significantly increased and correlated with the amplification of SPHK1 in DFSP samples (Figure 2e). These results were further verified through IHC experiments (Figure 2e).

### Structure variants identified in dermatofibrosarcoma protuberans

Compared with whole-exome sequencing or array-based comparative genomic hybridization, WGS is able to comprehensively reveal large-scale structural rearrangements. Here, eight structure variant (SV) subtypes were obtained in the 52 DFSP samples, i.e. deletion, inversion, tandem duplication, amplified inversion, foldback inversion, unbalanced inversion, intrachromosomal or interchromosomal translocation; the gene list corresponding to each SV is presented in Table S6. The number of SVs ranged from nine to 1301, with a median of 312.5 (Figure 3a, top panel). Among those SV subtypes, inversion, intrachromosomal and interchromosomal translocation accounted for the majority, and were dominant in 23%, 20% and 45%, respectively (Figure 3a, top panel). SVs are known to regulate oncogenes or tumour suppressor gene expression, which directly contributes to carcinogenesis,<sup>19,20</sup> as shown in Figure 3a (bottom panel). We also found a number of key genes in the signal transduction pathways that present SVs, including genes that were observed to be significantly altered in our



**Figure 2** Copy-number variants and experimental verification in dermatofibrosarcoma protuberans (DFSPs). (a) The raw copy number of autosomes in each sample of the cohort. The blocks in red and in blue denote copy-number amplification and deletion, respectively. (b) The identified significant amplification segments (red) and deletion segments (blue), and the corresponding oncogene or suppressor genes. (c) Comparison of tumour size, Ki67 expression, and duration between 17q25-amplification and wildtype samples. Significance was determined by using two-tailed Wilcoxon rank sum test ( $*P < 0.05$ ). (d) Verification and comparison of copy number of genes *SPHK1* and *ITGB4* between 17q25-amplification (positive) and wildtype samples (negative). Copy-number variation of genes *SPHK1* and *ITGB4* was explored by quantitative polymerase chain reaction (qPCR), which was performed three times for each DNA sample. Significance was evaluated using two-tailed Mann–Whitney U-test ( $***P < 0.001$ ,  $****P < 0.0001$ ). (e, f) Immunohistochemistry staining of *SPHK1* (1 : 200) and p-STAT3 (1 : 200) in *SPHK1*-amplification (positive) and wildtype samples (negative) DFSP tissues as described in the Materials and methods section. Representative images (e) and bar chart graphs of *SPHK1*/p-STAT3 staining intensity (f) were taken. Significance was evaluated using two-tailed Mann–Whitney U-test ( $**P < 0.01$ ).

previous analyses, such as *KMT2C*, *BRCA1*, *SPHK1*, *CDKN2A/B*, *COL1A1* and *PDGF $\beta$* .

In addition, clustering of breakpoint was applied to show complex genomic rearrangements in DFSP, and chromosomes 17 and 22 were found to have higher variation rates, which was consistent with the previous CNV analysis (Figure 3b). Notably, we evaluated clinical features significantly related to chromosomes 17 and 22 (Figure 3c, d) and found that the tumour size of the samples with chromosome 17 and 22 abnormalities was significantly larger than that of samples without variations. These findings indicate that DFSP is characterized by an abnormal genomic profile, particularly in chromosomes 17 and 22, which may play a key role in the tumorigenesis of DFSP.

### Fusion genes identified in dermatofibrosarcoma protuberans

Previous studies proved that the fusion of *COL1A1*-*PDGF $\beta$*  [t(17;22)] is one of the remarkable features in DFSP. Therefore, FISH analysis was applied to 43 samples, showing that the rate of *COL1A1*-*PDGF $\beta$*  fusion reached 74% (Figure 4a, top panel). Images of samples are presented in Figure S8 (see Supporting Information) and a representative sample is presented in Figure 4b. In addition, based on the WGS files, BreakID was also applied to identify fusion events and the result showed that the fusion frequency of *COL1A1*-*PDGF $\beta$*  was 72% in 43 samples. Compared with the FISH analysis results for those samples, the prediction coincidence rate reached 84% (Figure 4a, Table S7b; see Supporting Information). Interestingly, we discovered that the breakpoint of *PDGF $\beta$*  stably occurred between exon 1 and exon 2 of transcript variant 1 (NM\_002608) or transcript variant 2 (NM\_033016), while *COL1A1* randomly broke between exon 5 and exon 48 of transcript (NM\_000088) (Table S7a). The *COL1A1*-*PDGF $\beta$*  fusion in the WGS binary alignment and map (BAM) files of DFSP samples were further confirmed through the Integrative Genomics Viewer (IGV) (Figure S9a; see Supporting Information). As expected, those samples with *COL1A1*-*PDGF $\beta$*  fusion presented significantly higher Ki67 expression (Figure 4c).

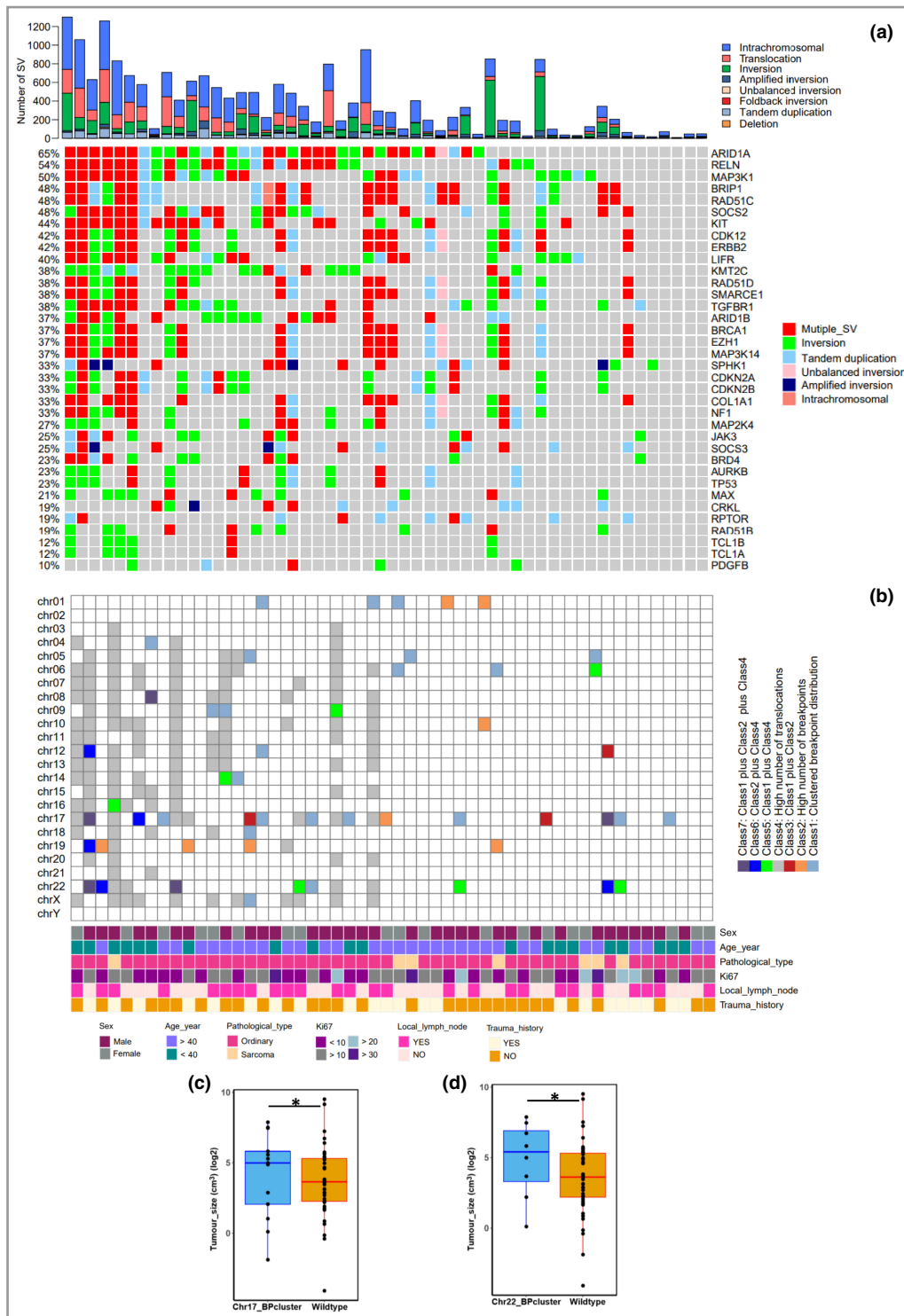
Moreover, as shown in Figure 4a (bottom panel), some fusion events were also detected by BreakID (Table S7b and Table S8; see Supporting Information). In particular, the frequency of *SLC2A5*-*BTBD7* [t(1;14)] exceeded 42%. Similarly, we discovered that the breakpoint of *SLC2A5* stably occurred

between exon 1 and exon 2 of transcript variant 2 (NM\_001135585), while *BTBD7* alternatively occurred at exon 7 of transcript variant 3 (NM\_001289133) or between exon 7 and exon 8 of transcript variant 3 (NM\_001289133) (Table S7a). Furthermore, the *SLC2A5*-*BTBD7* fusion in the WGS BAM files of DFSP samples was further confirmed through IGV (Figure S9b). We then performed PCR to verify the sequence before and after the fusion breakpoint. The sequence around the breakpoint in the fusion variant is shown in Figure 4d, demonstrating that the fusion occurred.

### Discussion

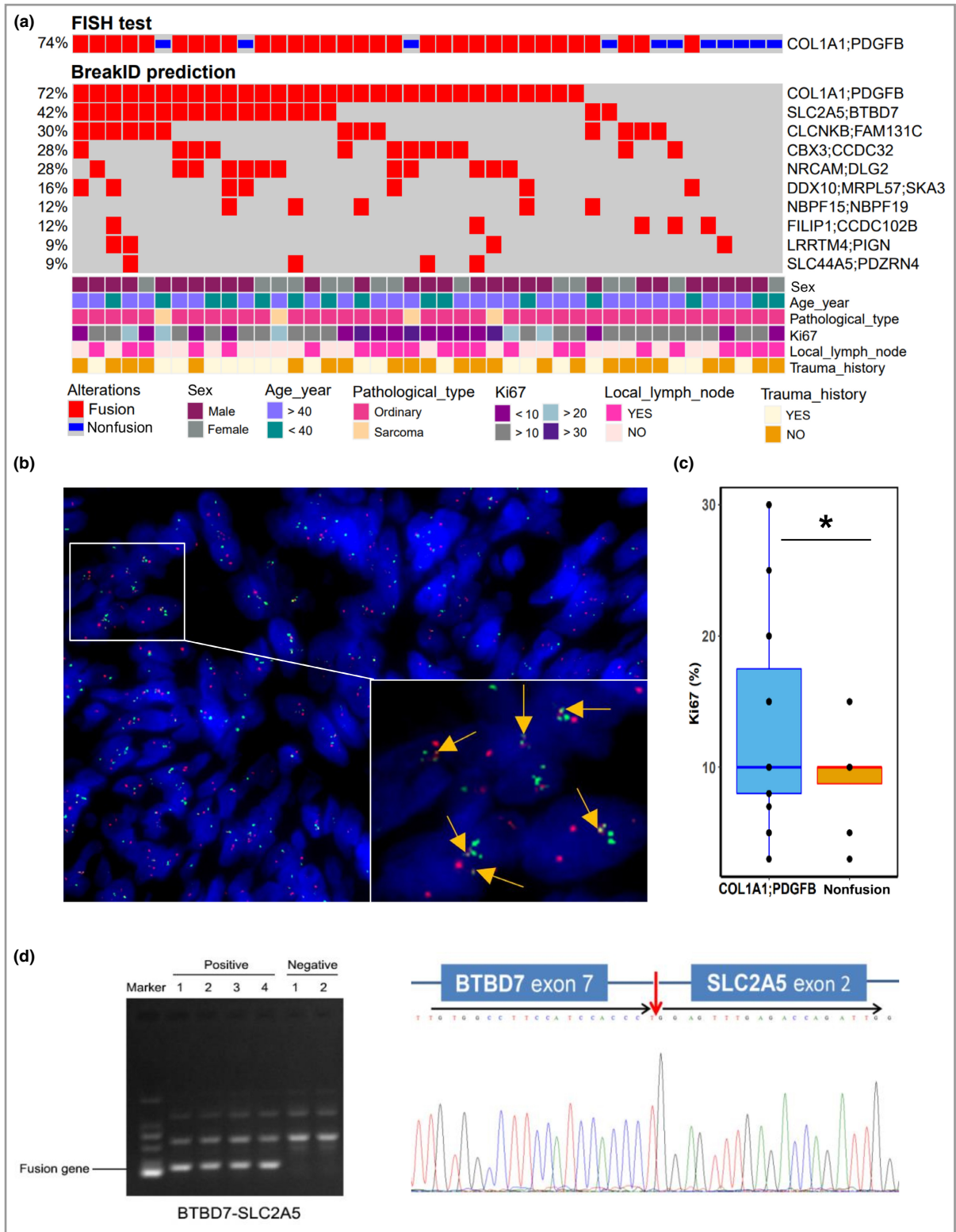
In this study, the landscape of genomic alterations for each DFSP revealed by WGS was summarized separately (Figure S10; see Supporting Information). Primarily, a high frequency of C > T and T > G mutations and a low TMB were detected in DFSP (Figure S1 and Figure S3). Mutational signature 1 is a major mutation type in DFSP, which is characterized by C > T mutations at CpG dinucleotides caused by deaminated 5 mC at CpG sites that have not been repaired before DNA replication, resulting in higher mutations in replication rounds.<sup>21,22</sup> Therefore, signature 1, which is described as a 'clock-like' signature, is believed to be induced with mutation accumulation during DNA replication.<sup>23</sup> Moreover, mutational signature 6 was also observed in DFSP, featuring a C > T transition at NpCpG mutation sites. *BRCA1* is well known to play an essential role in DNA damage repair, and loss of function of *BRCA1* triggers nonhomologous end-joining, resulting in microhomology-mediated double-strand break repair.<sup>24–26</sup> DNA repair is an important signalling pathway in DFSP (Figure 5), in which *BRCA1* is a key node with high frequency of SV (37%), amplification (81%) and mutation (6%), suggesting that cellular abnormalities in the DNA repair process play essential roles in the pathogenesis of DFSP.

TMB constitutes a quantitative analysis for measuring the number of somatic mutations per megabase in tumour cells, and the average TMB in different tumours is highly variable, ranging from about 0.5 to 12.<sup>27,28</sup> Comparatively, DFSP presented a relatively low TMB with a median of 3.49 (Figure S3). A higher TMB is reported to be a good prognostic marker, especially for immune checkpoint inhibitor outcomes,<sup>29,30</sup> and our finding showed that the TMB in patients

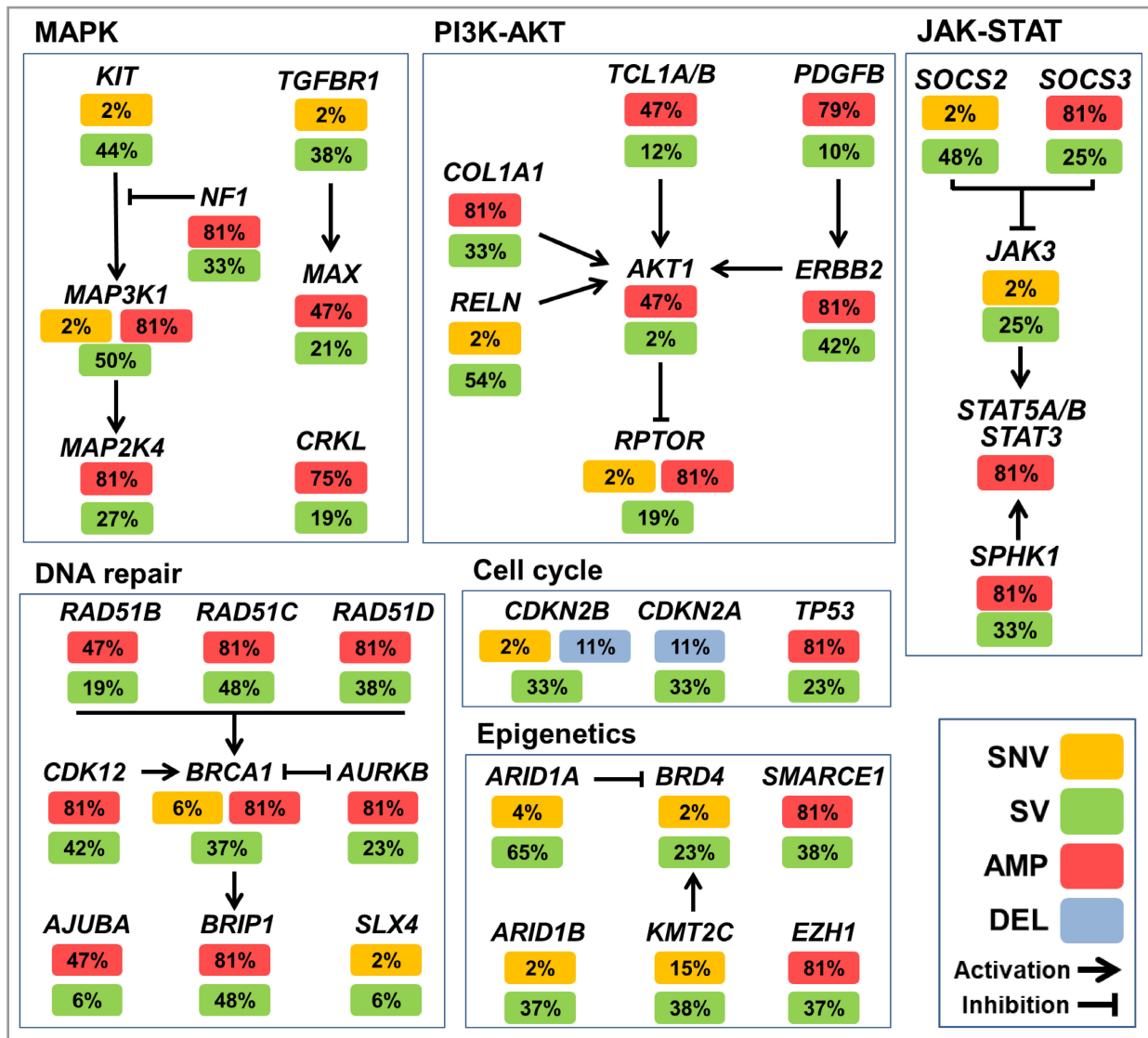


**Figure 3** Structure variants (SVs) identified in dermatofibrosarcoma protuberans (DFSP). (a) Summary of SVs in each sample of the cohort (top). Eight SV subtypes were identified in the DFSP samples ( $n = 52$ ) by Delly, and mutational waterfall of some key genes in the signal transduction pathways (bottom). (b) Detection of the clustering of breakpoint based on the complex genomic rearrangements as described in the Materials and methods section. (c, d). Comparison of tumour size between the Chr17-BPcluster (chromosomes 17) and wildtype samples (c), and comparison of tumour size between the Chr22-BPcluster (chromosomes 22) and wildtype samples (d). Significance was evaluated using the two-tailed Mann–Whitney U-test (\*\* $P < 0.01$ ).





**Figure 4** Fusion genes identified in dermatofibrosarcoma protuberans. (a) The COL1A1-PDGFB fusion tested by fluorescence in situ hybridization (FISH) in the 43 remaining subset samples (top) and those fusion genes identified by BreakID (bottom). (b) FISH analysis of COL1A1-PDGFB fusion in T29 sample (COL1A1, green; PDGFβ red). (c) Comparison of Ki67 expression between those cases with and without COL1A1-PDGFB fusion (two-tailed t-test). (d) Validation of fusion BTBD7-SLC2A5 by polymerase chain reaction (PCR) amplification (T27, T29, T32, T36, T28, T37). The sequences before and after the fusion breakpoint were first amplified using PCR, and the PCR products were analysed by electrophoresis on 1.0% agarose gels. The sequences were applied to the Basic Local Alignment Search Tool for the corresponding gene on the National Center for Biotechnology Information database.



**Figure 5** Summary of genomic alterations of several key signal transduction pathways in dermatofibrosarcoma protuberans (DFSP). DNA repair, cell cycle, mitogen-activated protein kinase (MAPK), phosphoinositide 3-kinase (PI3K)-AKT and Janus kinase (JAK)-signal transducer and activator of transcription (STAT) pathways were found to be pivotal pathways in DFSP tumorigenesis. SNV, single-nucleotide variants; SV, structure variants; AMP, amplification; DEL, deletion.

with recurrent DFSP was significantly lower than that in those with nonrecurrent DFSP (Figure 1c), indicating that a low TMB might be a potential predictive marker for relapse of DFSP.

MUC family genes, including MUC6 (25%) and MUC4 (21%), were significantly associated with mutations in DFSP. Interestingly, mutations of MUC6 and MUC4 were mutually exclusive (Figure S4), thus DFSP was classified into three subtypes based on mutation profiles (Figure 1b, bottom panel). Furthermore, KMT2C (15%), HERC2 (11%), PRSS1 (8%) and BRCA1 (6%) were identified to be related to tumorigenesis of DFSP (Figure 1b). KMT2C (known as MLL3) is a tumour suppressor gene characterized by histone methyltransferase activity to maintain H3K4 monomethylation at enhancers,<sup>31–33</sup> and is one of the most frequently mutated genes in various

tumours,<sup>33–35</sup> and the loss of KMT2C facilitates cellular transformation.<sup>36,37</sup>

Our findings showed that several genomic rearrangement events occur in various chromosomal regions, particularly in 17q25 and 22q13, which causes amplification of oncogenes including AKT1, SPHK1, COL1A1 and PDGF $\beta$ . Therefore, as a consequence of genomic variation events benefiting tumour cells, proliferation advantage may play an important role for driving DFSP. We also noticed the complexity and diversity of the SVs in DFSP, which are involved in well-known tumour suppressor or oncogene genes, such as CDKN2A/B, BRCA1, SPHK1 and COL1A1, indicating that suppression of those pathways might be a novel potential strategy for DFSP treatment.

Given the high frequency of COL1A1-PDGFB fusion in DFSP, COL1A1 and PDGF $\beta$  were considered to be oncogenic drivers



in DFSP. Our findings revealed that these two genes also exhibited high frequency of amplification and SVs, indicating that the variation of genome structure is the main cause of the abnormality of these two genes. In this study, the rate of COL1A1-PDGFB fusion is 74%, which is lower than the rate found in previous studies.<sup>6,38–41</sup> Therefore, we considered that there could be two possible explanations; one relating to the patients' ethnicity, and another related to age. We compared our findings with different DFSP studies on the relationship between patients with COL1A1-PDGFB fusion and age, as shown in Figure S11a (see Supporting Information). It appeared that old age was positively correlated with the rate of positive COL1A1-PDGFB fusion. Furthermore, our findings also demonstrated that disease duration with confirmed COL1A1-PDGFB fusion was significantly longer than for patients who were negative for COL1A1-PDGFB fusion (Figure S11b), indicating that COL1A1-PDGFB fusion could be related to the course of disease.

In addition to the COL1A1-PDGFB fusion gene, we identified another high-frequency fusion gene in DFSP, namely SLC2A5-BTBD7 (Figure 4), which has been reported to be a fusion gene in prostate cancer and acute lymphoblastic leukaemia.<sup>42,43</sup> SLC2A5 encodes Glut5, which is one of the facilitative glucose transporters that responds to fructose transport.<sup>44–46</sup> Evidence showed that inhibition of Glut5 expression significantly attenuates a malignant phenotype in breast cancer cells and Glut5 may act as a novel biomarker for early tumour diagnosis.<sup>47,48</sup> BTBD7 belongs to the POZ gene family, which is characterized by the BTB/POZ protein–protein interaction motif.<sup>49</sup> BTBD7 plays a critical role in tumour cell epithelial mesenchymal transition through Snail2, fibronectin and E-cadherin in lung cancer and hepatocellular carcinoma;<sup>50,51</sup> however, the function of SLC2A5-BTBD7 fusion remains to be further clarified.

This is the first study to describe comprehensive genomic alterations in DFSP. Through WGS, we identified gene mutations, CNVs and fusion gene profiles in DFSP. DNA repair, cell cycle, mitogen-activated protein kinase, phosphoinositide 3-kinase-AKT and Janus kinase-signal transducer and activator of transcription pathways are pivotal pathways involving the altered genes, which may shed light on the identification of potential diagnostic and therapeutic targets in DFSP.

## Acknowledgments

We acknowledge BioGenius for running WGS raw data analysis pipeline (<https://www.biogenius.cn/>), and Yiwang Zhang, Haifeng Li for their work in the histopathological diagnosis of DFSP.

## References

- Kreicher KL, Kurlander DE, Gittleman HR *et al.* Incidence and survival of primary dermatofibrosarcoma protuberans in the United States. *Dermatol Surg* 2016; **42** (Suppl. 1):S24–31.
- Sirvent N, Maire G and Pedeutour F. Genetics of dermatofibrosarcoma protuberans family of tumors: from ring chromosomes to tyrosine kinase inhibitor treatment. *Genes Chromosomes Cancer* 2003; **37**:1–19.
- Simon MP, Pedeutour F, Sirvent N *et al.* Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma. *Nat Genet* 1997; **15**:95–8.
- Pedeutour F, Simon MP, Minoletti F *et al.* Ring 22 chromosomes in dermatofibrosarcoma protuberans are low-level amplifiers of chromosome 17 and 22 sequences. *Cancer Res* 1995; **55**:2400–3.
- Patel KU, Szabo SS, Hernandez VS *et al.* Dermatofibrosarcoma protuberans COL1A1-PDGFB fusion is identified in virtually all dermatofibrosarcoma protuberans cases when investigated by newly developed multiplex reverse transcription polymerase chain reaction and fluorescence in situ hybridization assays. *Hum Pathol* 2008; **39**:184–93.
- Salgado R, Llombart B, R MP *et al.* Molecular diagnosis of dermatofibrosarcoma protuberans: a comparison between reverse transcriptase-polymerase chain reaction and fluorescence in situ hybridization methodologies. *Genes Chromosomes Cancer* 2011; **50**:510–17.
- Heller S, Scheibenpflug L, Westermark B, Nistér M. PDGF B mRNA variants in human tumors with similarity to the v-sis oncogene: expression of cellular PDGF B protein is associated with exon 1 divergence, but not with a 3'UTR splice variant. *Int J Cancer* 2000; **85**:211–22.
- Sinovic J, Bridge JA. Translocation (2;17) in recurrent dermatofibrosarcoma protuberans. *Cancer Genet Cytogenet* 1994; **75**:156–7.
- Pedeutour F, Simon MP, Minoletti F *et al.* Translocation, t(17;22)(q22;q13), in dermatofibrosarcoma protuberans: a new tumor-associated chromosome rearrangement. *Cytogenet Cell Genet* 1996; **72**:171–4.
- Sonobe H, Furihata M, Iwata J *et al.* Dermatofibrosarcoma protuberans harboring t(9;22)(q32;q12.2). *Cancer Genet Cytogenet* 1999; **110**:14–18.
- Bianchini L, Maire G, Guillot B *et al.* Complex t(5;8) involving the CSPG2 and PTK2B genes in a case of dermatofibrosarcoma protuberans without the COL1A1-PDGFB fusion. *Virchows Arch* 2008; **452**:689–96.
- Chakravarty D, Gao J, Phillips S *et al.* OncoKB: a precision oncology knowledge base. *JCO Precision Oncology* 2017; **2017**:PO.17.00011.
- Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriverCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* 2013; **29**:2238–44.
- Miao D, Margolis CA, Vokes NI *et al.* Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet* 2018; **50**:1271–81.
- Yang Y, Zhang J, Chen Y *et al.* MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pancreatic cancer. *Clin Transl Med* 2020; **10**:e155.
- Gautam SK, Kumar S, Dam V *et al.* MUCIN-4 (MUC4) is a novel tumor antigen in pancreatic cancer immunotherapy. *Semin Immunol* 2020; **47**:101391.
- Favero F, Joshi T, Marquard AM *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 2015; **26**:64–70.
- Mermel CH, Schumacher SE, Hill B *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12**:R41.
- Li Y, Roberts ND, Wala JA *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* 2020; **578**:112–21.

- 20 Campbell PJ, Yachida S, Mudie LJ *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 2010; **467**:1109–13.
- 21 Pfeifer GP. Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* 2006; **301**:259–81.
- 22 Aparicio SAJR, Alexandrov LB, Nik-Zainal S *et al.* Signatures of mutational processes in human cancer. *Nature* 2013; **500**:415–21.
- 23 Alexandrov LB, Jones PH, Wedge DC *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet* 2015; **47**:1402–7.
- 24 Farmer H, McCabe N, Lord CJ *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; **434**:917–21.
- 25 Bau DT, Fu YP, Chen ST *et al.* Breast cancer risk and the DNA double-strand break end-joining capacity of nonhomologous end-joining genes are affected by BRCA1. *Cancer Res* 2004; **64**:5013–19.
- 26 Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci* 2004; **95**: 866–71.
- 27 Chalmers ZR, Connelly CF, Fabrizio D *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* 2017; **9**:34.
- 28 Zehir A, Benayed R, Shah RH *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017; **23**:703–13.
- 29 Chan TA, Yarchoan M, Jaffee E *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* 2019; **30**: 44–56.
- 30 Lee M, Samstein RM, Valero C *et al.* Tumor mutational burden as a predictive biomarker for checkpoint inhibitor immunotherapy. *Hum Vaccin Immunother* 2020; **16**:112–15.
- 31 Herz HM, Garruss A and Shilatifard A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci* 2013; **38**:621–39.
- 32 Herz HM, Mohan M, Garruss AS *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev* 2012; **26**:2604–20.
- 33 Fagan RJ and Dingwall AK. COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Lett* 2019; **458**:56–65.
- 34 Kandoth C, McLellan MD, Vandin F *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502**:333–9.
- 35 Lawrence MS, Stojanov P, Mermel CH *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; **505**:495–501.
- 36 Cho SJ, Yoon C, Lee JH *et al.* KMT2C mutations in diffuse-type gastric adenocarcinoma promote epithelial-to-mesenchymal transition. *Clin Cancer Res* 2018; **24**:6556–69.
- 37 MacFawn I, Wilson H, Selth LA *et al.* Grainyhead-like-2 confers NK-sensitivity through interactions with epigenetic modifiers. *Mol Immunol* 2019; **105**:137–49.
- 38 Stacchiotti S, Pantaleo MA, Negri T *et al.* Efficacy and biological activity of imatinib in metastatic dermatofibrosarcoma protuberans (DFSP). *Clin Cancer Res* 2016; **22**:837–46.
- 39 McArthur GA, Demetri GD, van Oosterom A *et al.* Molecular and clinical analysis of locally advanced dermatofibrosarcoma protuberans treated with imatinib: Imatinib Target Exploration Consortium Study B2225. *J Clin Oncol* 2005; **23**:866–73.
- 40 Kérob D, Porcher R, Vérola O *et al.* Imatinib mesylate as a preoperative therapy in dermatofibrosarcoma: results of a multicenter phase II study on 25 patients. *Clin Cancer Res* 2010; **16**:3288–95.
- 41 Yang JH, Hu WH, Li F *et al.* [Detection of COL1A1/PDGFB fusion transcripts in dermatofibrosarcoma protuberans by reverse transcriptase-polymerase chain reaction using paraffin-embedded tissues]. *Zhonghua Bing Li Xue Za Zhi* 2003; **32**:409–12 (in Chinese).
- 42 Zeng T, Fedeli MA, Tanda F *et al.* Whole-exome sequencing of prostate cancer in sardinian identify recurrent UDP-glucuronosyltransferase amplifications. *J Cancer* 2021; **12**:438–50.
- 43 Zhang J, Loh ML, Ma X *et al.* Comparison of mutational profiles of diagnosis and relapsed pediatric B-acute lymphoblastic leukemia: a report from the COG ALL target project. *Blood* 2013; **122**:824.
- 44 Burant CF, Takeda J, Brot-Laroche E *et al.* Fructose transporter in human spermatozoa and small intestine is GLUT5. *J Biol Chem* 1992; **267**:14523–6.
- 45 Blakemore SJ, Aledo JC, James J *et al.* The GLUT5 hexose transporter is also localized to the basolateral membrane of the human jejunum. *Biochem J* 1995; **309**:7–12.
- 46 Nomura N, Verdon G, Kang HJ *et al.* Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* 2015; **526**:397–401.
- 47 Zamora-León SP, Golde DW, Concha, II *et al.* Expression of the fructose transporter GLUT5 in human breast cancer. *Proc Natl Acad Sci USA* 1996; **93**: 1847–52.
- 48 Hamann I, Krys D, Glubrecht D *et al.* Expression and function of hexose transporters GLUT1, GLUT2, and GLUT5 in breast cancer-effects of hypoxia. *FASEB J* 2018; **32**:5104–18.
- 49 Onodera T, Sakai T, Hsu JC *et al.* Btd7 regulates epithelial cell dynamics and branching morphogenesis. *Science* 2010; **329**:562–5.
- 50 Tao YM, Huang JL, Zeng S *et al.* BTB/POZ domain-containing protein 7: epithelial-mesenchymal transition promoter and prognostic biomarker of hepatocellular carcinoma. *Hepatology* 2013; **57**:2326–37.
- 51 Shu J, Wang L, Han F *et al.* BTBD7 downregulates e-cadherin and promotes epithelial-mesenchymal transition in lung cancer. *Biomed Res Int* 2019; **2019**:5937635.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Figure S1** Somatic alterations in dermatofibrosarcoma protuberans.

**Figure S2** Single-nucleotide substitutions identified in dermatofibrosarcoma protuberans.

**Figure S3** Comparison of somatic mutations between dermatofibrosarcoma protuberans and the 33 tumours of The Cancer Genome Atlas database.

**Figure S4** Mutual interaction of driver genes identified in dermatofibrosarcoma protuberans.

**Figure S5** Gene ontology (GO) enrichment analysis. (a–c) The identified somatic mutation genes in dermatofibrosarcoma protuberans underwent GO enrichment on (a) biological process, (b) molecular function and (c) cell component.

**Figure S6** Segments of copy-number variants in dermatofibrosarcoma protuberans.

**Figure S7** Gene ontology (GO) enrichment analysis. (a–c) The identified genes with copy-number amplification in dermatofibrosarcoma protuberans underwent GO enrichment on (a) biological process, (b) molecular function, and (c) cell component.

**Figure S8** Fluorescence in situ hybridization analysis of COL1A1-PDGFB fusion in 43 DFSP samples.

**Figure S9** The fusion sequence of (a) PDGFB-COL1A1 (b) and SLC2A5-BTBD7 were detected by Integrative Genomics Viewer in the whole-genome sequencing binary alignment and map files of T12 dermatofibrosarcoma protuberans case.

**Figure S10** The landscape of genomic alterations of 52 dermatofibrosarcoma protuberans revealed by whole-genome sequencing.

**Figure S11** The relationship between age, the course of disease and the frequency of COL1A1-PDGFB fusion.

**Table S1a** Patient clinical and demographic characteristics.

**Table S1b** The immunohistochemistry results for all patients.

**Table S2a** The identified somatic variants in dermatofibrosarcoma protuberans.

**Table S2b** The identified single-nucleotide substitutions in dermatofibrosarcoma protuberans.

**Table S2c** Mutational signatures identified in dermatofibrosarcoma protuberans.

**Table S2d** Tumour mutation burden (TMB) per megabase in 53 dermatofibrosarcoma protuberans samples.

**Table S2e** Somatic genes.

**Table S2f** The identified driver genes using oncodriveCLUST algorithm in dermatofibrosarcoma protuberans.

**Table S3a** Gene ontology enrichment analysis on biological process based on the somatic genes with high frequency (> 5%, at least three samples).

**Table S3b** Gene ontology enrichment analysis on molecular function based on the somatic genes with high frequency (> 5%, at least three samples).

**Table S3c** Gene ontology enrichment analysis on cell component based on the somatic genes with high frequency (> 5%, at least three samples).

**Table S4a** The significant amplification peaks and deletion peaks in dermatofibrosarcoma protuberans.

**Table S4b** The significant amplification peaks and the mapping genes in dermatofibrosarcoma protuberans.

**Table S4c** The significant deletion peaks and the mapping genes in dermatofibrosarcoma protuberans.

**Table S5a** Gene ontology enrichment analysis on biological process based on the genes mapped by significant amplification segments.

**Table S5b** Gene ontology enrichment analysis on molecular function based on the genes mapped by significant amplification segments.

**Table S5c** Gene ontology enrichment analysis on cell component based on the genes mapped by significant amplification segments.

**Table S6** Gene list corresponding to each structure variant.

**Table S7a** The identified fusion genes and their breakpoints information.

**Table S7b** The identified fusion genes and mutation samples in 43 dermatofibrosarcoma protuberans samples.

**Table S8** The fluorescence in situ hybridization results and BreakID prediction of COL1A1-PDGFB fusion gene in 43 dermatofibrosarcoma protuberans samples.

**Appendix S1** Whole-genome sequencing.