

ARTICLE

Positional cloning and next-generation sequencing identified a *TGM6* mutation in a large Chinese pedigree with acute myeloid leukaemia

Li-li Pan^{1,2,5}, Yuan-mao Huang^{2,5}, Min Wang², Xiao-e Zhuang², Dong-feng Luo², Shi-cheng Guo³, Zhi-shun Zhang⁴, Qing Huang², Sheng-long Lin² and Shao-yuan Wang^{1,2}

An inherited predisposition to acute myeloid leukaemia (AML) is exceedingly rare, but the investigation of these families will aid in the delineation of the underlying mechanisms of the more common, sporadic cases. Three AML predisposition genes, *RUNX1*, *CEBPA* and *GATA2*, have been recognised, but the culprit genes in the majority of AML pedigrees remain obscure. We applied a combined strategy of linkage analysis and next-generation sequencing (NGS) technology in an autosomal-dominant AML Chinese family with 11 cases in four generations. A genome-wide linkage scan using a 500K SNP genotyping array was conducted to identify a previously unreported candidate region on 20p13 with a maximum multipoint heterogeneity LOD (HLOD) score of 3.56 ($P=0.00005$). Targeted NGS within this region and whole-exome sequencing (WES) revealed a missense mutation in *TGM6* (RefSeq, NM_198994.2:c.1550T>G, p.(L517W)), which cosegregated with the phenotype in this family, and was absent in 530 healthy controls. The mutated amino acid was located in a highly conserved position, which may be deleterious and affect the activation of *TGM6*. Our results strongly support the candidacy of *TGM6* as a novel familial AML-associated gene.

European Journal of Human Genetics (2015) 23, 218–223; doi:10.1038/ejhg.2014.67; published online 23 April 2014

INTRODUCTION

Familial acute myeloid leukaemia (AML) pedigrees have been rarely reported, partially due to the high mortality of leukaemia, significant variability in the age at onset, and the small size of modern families. These factors complicate researches on AML predisposition families. Three genes have been causally linked to leukaemia susceptibility: *RUNX1*, *CEBPA* and *GATA2*.^{1–3} In addition, *TERC* and *TERT* mutations have also been reported in patients of familial MDS/AML.⁴ However, familial AML is a genetically heterogeneous disorder, and the culprit genes in the majority of AML pedigrees remain obscure.⁵ Moreover, accumulating experimental and epidemiological evidence suggests that no single mutation is sufficient to produce AML.⁶ Thus, additional pedigrees are required to fully delineate the underlying mechanism of leukaemia.

We report on an unusual AML-predisposed family with 11 cases in four generations, which is the second largest number of AML cases in a single family in the world. We identified a missense mutation in *TGM6* in this family that cosegregated with the phenotype and predicted functional damage using a combined strategy of linkage analysis and next-generation sequencing (NSG).

MATERIALS AND METHODS

Patients and materials

This study was approved by the Expert Committee of Fujian Medical University Union Hospital in China (equivalent to an institutional review

board). All participants provided written informed consent before enrolment. One family (Figure 1) with 40 members, including 11 AML patients in four consecutive generations (Supplementary Table 1), was closely followed for several years. Seven patients had been reported by He *et al*⁷ in 1994, and we identified another four newly diagnosed cases, IV-19, IV-14, III-15 and I-3, through a follow-up investigation of this family. All patients were diagnosed according to FAB classifications. The criteria of ‘members at potential preleukaemic phase’ were defined in Supplementary Material. Blood samples were obtained from 2 patients (III-15 and IV-19), 6 members at potential preleukaemic phase (Table 1), 11 unaffected family members and 2 spouses. Samples from 21 family members were not obtained for that 8 members died in childhood from unknown causes, 4 members refused to be investigated and 9 patients died with AML before samples were obtained for research. Healthy individuals ($n=530$) of matched geographic ancestry were included. Patients III-15 and IV-19 have been excluded of copy number variants (unpublished data) and known causative variants in *CEBPA*, *RUNX1* and *GATA2* via sequence analysis.

Genotyping, linkage and haplotype analysis

Thirteen members of the family (Figure 1) were genotyped using the Affymetrix GeneChip Human Mapping Array 500K set according to the manufacturer’s recommended protocol (Affymetrix, Santa Clara, CA, USA).⁸ MERLIN (v1.1.2)⁹ were used to perform multipoint linkage analysis under the non-parametric model and dominant model (Supplementary Material). SNP markers within the high-score regions were chosen to construct the haplotype by the GENEHUNTER program v2.1.¹⁰

¹Fujian Institute of Hematology, Fujian Provincial Key Laboratory on Hematology, Department of Hematology, Fujian Medical University Union Hospital, Fuzhou, China; ²Union Clinical Medical College, Fujian Medical University, Fuzhou, China; ³State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; ⁴State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

*Correspondence: Professor S-y Wang, Fujian Institute of Hematology, Fujian Provincial Key Laboratory on Hematology, Department of Hematology, Fujian Medical University Union Hospital, 29 Xinquan Road, Fuzhou 350001, China. Tel: +86 591 83357896; Fax: +86 591 88599539; E-mail: mdsy.wang@gmail.com

⁵These authors contributed equally to this work and should be considered as co-first authors.

Received 14 March 2013; revised 16 January 2014; accepted 18 March 2014; published online 23 April 2014

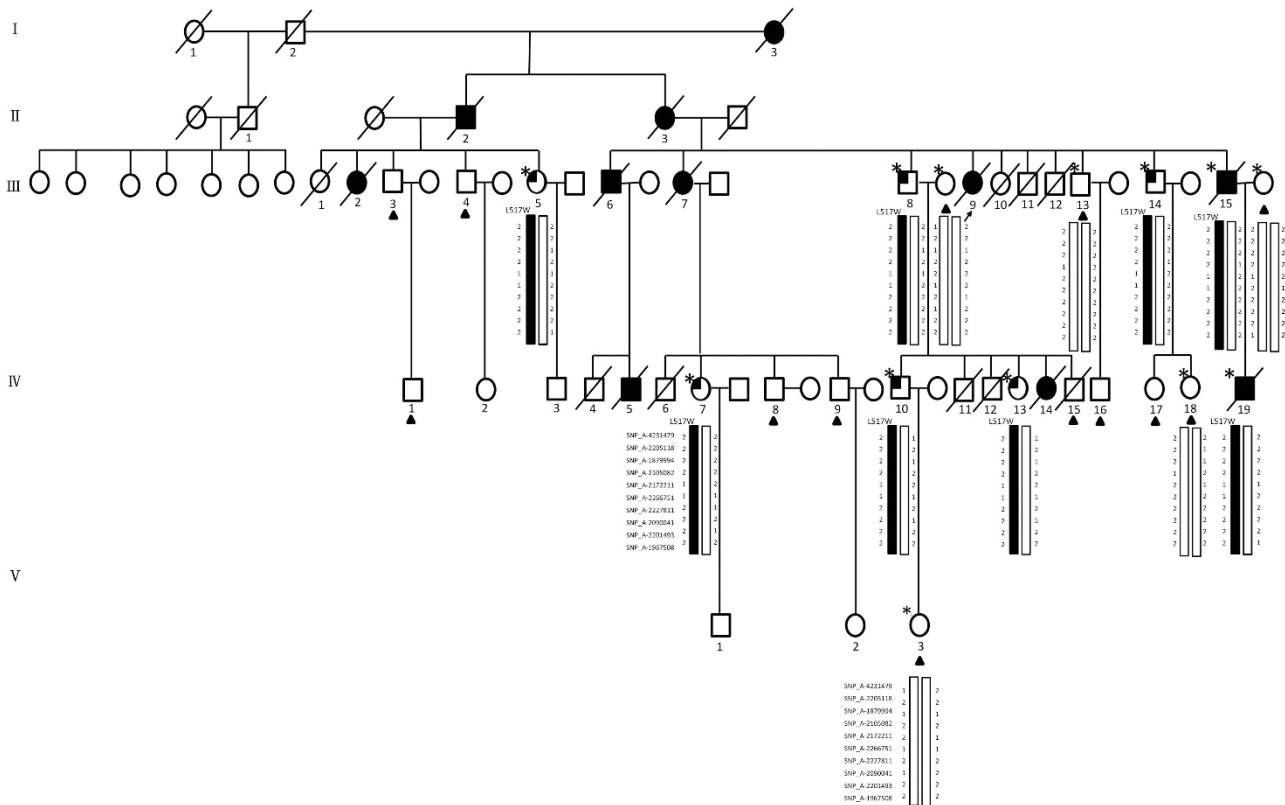


Figure 1 Familial AML pedigree. Squares represent males; circles represent females. The proband is indicated by an arrow. Deceased subjects are indicated by a diagonal line. Roman numerals denote generations. Arabic numerals indicate the position within the same generation. Open symbols represent unaffected individuals; closed symbols represent affected individuals. Partially filled symbols denote subjects at potential preleukaemic phase (Supplementary Material). Closed symbols with a diagonal line indicate that the individual died from AML. L517W represents subjects with the *TGM6* missense mutation, including two patients and six members at potential preleukaemic phase. Triangles represent subjects with a *TGM6* wild-type allele, including 11 unaffected family members and 2 spouses. Genotyped individuals are indicated by an asterisk, including 2 patients, 6 members at potential preleukaemic phases, 11 unaffected family members and 2 spouses. The most likely haplotype for the linked chromosome 20 markers is shown below each genotyped individual: black symbols denote the disease haplotype, and white symbols represent a normal haplotype. The order of markers from top to bottom is SNP_A-4231479, SNP_A-2205118, SNP_A-1879994, SNP_A-2105082, SNP_A-2172211, SNP_A-2266751, SNP_A-2227811, SNP_A-2090041, SNP_A-2201493 and SNP_A-1967508. The markers spanned a genomic distance from chr20:2337295 to chr20:4544193, and *TGM6* located within this region. Members who carried the disease haplotype also inherited the mutation in *TGM6*, which in turn prove the existence of the haplotype in this family.

Targeted NGS and WES

NimbleGen 385K microarrays were produced to capture the critical region at 20p13 (7.8–13 cM) in two patients (III-15 and IV-19). Libraries construction and sequencing were completed according to the manufacturer's instructions (Roche 454 company, Branford, CT, USA).¹¹ Sequence data were initially mapped to a human genome reference sequence (hg19) and annotated using the GS Reference Mapper software package (Roche). All variants were identified using ALLDiff and more stringent HCDiff approaches.¹² SNPs in the dbSNP138 and 1000 genome project databases (2013) were removed. The remaining HCDiff variants that were shared by the two samples were selected. We further explored the potential effects of these mutations using SIFT,¹³ Polyphen software¹⁴ and phyloP conservation score. We also performed whole-exome sequencing (WES) on the two patients plus a healthy family member following the manufacturer's instructions (Illumina, San Diego, CA, USA),¹⁵ and an in-house bioinformatics pipeline similar to targeted NGS was used for WES data analysis (Supplementary Material).

Mutation screening and molecular modelling

Sanger sequencing were performed to confirm the candidate coding variants that were identified by the targeted NGS and WES. The *TGM6* mutation was screened in 13 genotyped members, 8 unaffected family members and 530 healthy controls. The 3D molecular models of *TGM6* were built using homology modelling (Supplementary Material).

RESULTS

Linkage and haplotype analysis

A total of 13 Affymetrix Mapping 500K arrays were processed, which resulted in the generation of >6.5 million genotypes. The average SNP call rate and heterozygosity for the 13 genotyped individuals were 96.42% (93–98.34%) and 23.68% (24.06–24.42%), respectively. A total of 6480 SNP markers remained for the final linkage analysis after stringent tag SNP selection. The average information content for each of the 23 chromosomes was ranged from 0.803 to 0.912.

Multipoint analysis using the dominant and non-parametric models resulted in two potential linkage regions on 20p13 (maximum multipoint heterogeneity LOD (HLOD) = 3.56, $P = 0.00005$; non-parametric linkage (NPL) = 2.69, $P = 0.0002$, $Z = 16.27$; Figure 2 and Supplementary Table 5) and 18q22.1–22.3 (maximum HLOD = 1.57, $P = 0.007$; NPL = 1.28, $P = 0.008$, $Z = 3.24$; Supplementary Figure 1). A broad region on chromosome 20 extending from 7.84 to 13.04 cM (2162598–4475430) was associated with an average HLOD score of 3.42 (average $P = 0.00009$) and an average NPL score of 2.59 (average $P = 0.0003$) within the same region. Another region on chromosome 18, extending from 91.46 to 97.06 cM (66127086–69342671), was associated with an average HLOD score of 1.56 (average $P = 0.0074$)

Table 1 Clinical presentation of genotyped members

ID	Sex	Blood count (median)			Bone marrow aspirate	Karyotype	Clinical features	Vital status			
		WBC, g/l	Plt, g/dl	Hb, g/dl				(age at last follow-up, years)	Status in linkage analysis	TGM6 mutation	
IV19	M	13	58	20	63	Hypercellular, with 78% myeloperoxidase-positive blast cells, diagnosed with AML-M2 in February 2003.	Unidentified	He was treated with daunorubicin and cytarabine and entered a complete remission. He had a relapse within 3 months, never entered a second remission and died in November 2003.	Dead (13)	Affected	c.1550T>G
III15	M	37	69	45	65	Hypercellular, with 91% myeloperoxidase-positive blast cells, diagnosed with AML-M1 in March 2008.	2007: 46,XY; 2008 (after onset): 45,X,-Y.	He rejected chemotherapy for financial reasons, and died within 2 months.	Dead (37)	Affected	c.1550T>G
III5	F	43	5.1	213	85	Hypercellular; 3% erythroid precursors with dysplastic features, including megaloblastic changes, and mature red cells with anisocytosis.	The frequency of chromosome breakage events was increased 18% by DNA cross-linking agents exposure.	A long history of paleness and easy fatigue, and can not be explained by any underlying medical condition.	Alive (47)	Affected	c.1550T>G
III14	M	44	4.5	72	102	Hypercellular; 4.5% megakaryocytes with dysplastic features, including micromegakaryocytes and multinucleated megakaryocytes.	The frequency of chromosome breakage events was increased 4% by DNA cross-linking agents exposure.	A long history of easy bruising and frequent epistaxis.	Alive (48)	Affected	c.1550T>G
III8	M	60	3	153	132	Normocellular, 3.5% granulocytes with dysplastic features, including agranular myelocytes, hypersegmented or polyploid neutrophils.	Unidentified.	He was an obligate carrier for that his daughter died with AML. He did not show any symptoms at last follow-up.	Alive (64)	Affected	c.1550T>G
IV10	M	35	5.2	127	138	Hypercellular, with 4% myeloblasts.	2007:46,XXY,+15,-19,-21,(2q1,18q-).2008:46,XY,del(11q22).	He did not show any symptoms.	Alive (39)	Affected	c.1550T>G
IV7	F	41	13.1	87	101	Hypercellular; 3% granulocytes with dysplastic features, including agranular myelocytes and hypersegmented neutrophils; 3% erythroid precursors with dysplastic features, including nuclear atypia and some megaloblastic changes.	The frequency of chromosome breakage events was increased 8% by DNA cross-linking agents exposure.	A long history of gingiva bleeding, increased menstrual and paleness. Her mother died of AML.	Alive (45)	Affected	c.1550T>G
IV13	F	26	14.1	224	80	Hypercellular; 3% erythroid precursors with dysplastic features.	46, XX, 19qh+.	A history of unexplained recurrent miscarriage.	Alive (30)	Affected	c.1550T>G
III13	M	50	N	N	N	N	46,XY.	None.	Alive (54)	Unaffected	Wild type
III8's wife	F	55	N	N	N	N	46,XX.	None.	Alive (59)	Unaffected	Wild type
V3	F	11	N	N	N	N	46,XX.	None.	Alive (15)	Unaffected	Wild type
III15's wife	F	35	N	N	N	N	46,XX.	None.	Alive (39)	Unaffected	Wild type
IV18	F	27	N	N	N	N	46,XX.	None.	Alive (31)	Unaffected	Wild type

Abbreviations: N, normal; None, no haematological manifestations. The cytopenia of family members had been pre-excluded of iron-deficiency anaemia and megaloblastic anaemia.

and an average NPL score of 1.27 (average $P = 0.008$). Markers within these two regions were used to construct haplotypes for all of the genotyped members. And all of the affected members shared the same disease haplotype in the 20p13 linkage region, whereas the unaffected family members did not exhibit this haplotype, which again support this region as a candidate region. Affected members did not carry the same haplotype on 18q22.1–22.3, which thus excluded this region.

Targeted capture and 454 sequencing

A total of 680 540 and 740 628 HQ reads (quality values $\geq Q20$) with an average length of 346–359 bp were produced. We demonstrated that 97.97–98.40% of the reads mapped to the human genome, and 78.2–87.5% mapped to the target region, which were included in our downstream analyses. The average sequencing depths across the targeted intervals ranged from 36.1 to 34.8*. The sequencing depth was $>5^*$ for 93–95% and $>10^*$ for 86.7–88.5% of the targeted regions. We finally generated a total of 17 829–17 849 variants in the target region, and 4166–4191 of these variants were high-confidence variants (HCDiffs). There were 36 HCDiffs shared by the two sequenced family patients after excluding synonymous and SNPs (Supplementary Table 2). Two of them were exonic variants, 15 intronic variants, 15 intergenic variants and 4 UTR variants (Supplementary Table 3). The exonic variants affected *TGM6* and *CENPB*, respectively.

Whole-exome sequencing

We generated an average of 57 712 134 reads per sample as paired-end, 90-bp reads; 52 299 061 reads (90.62%) passed the quality assessment and aligned to the human reference sequence; the coverage of target region was 98.87% and the average sequencing depth on target was 52.4-fold. The sequencing depth was $>4^*$ for 97.07% and 10^* for 92.77% of the targeted regions. We detected 122–126 variants per sample within the linkage region at 20p13 (Supplementary Table 2). After excluding SNPs in dbSNP138 or 1000 Genome Project

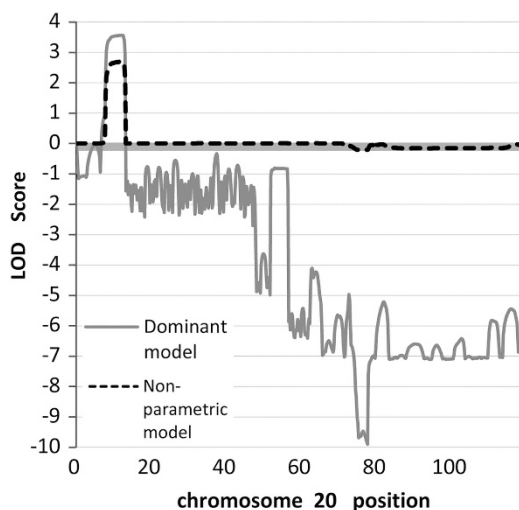


Figure 2 LOD plots for chromosome 20. X axis represent genetic distance (cM), and y axis represent the corresponding LOD score. LOD score peak located within a region ranging from 7.84 to 13.04 cM (chr20: 2162598–4475430) with an average HLOD score of 3.42, and *TGM6* was within this region. The LOD score was significantly higher in dominant model than non-parametric model, which support a dominant transmission of the disease in the family we studied.

databases (2013), only 8–12 variants remained and 4 of them were shared by the two family patients but absent in the unaffected member; 2 exonic variants and 2 intronic variants (Supplementary Table 4). The exonic variants affected *TGM6* and *SIRPA*, respectively.

TGM6 mutation screening and molecular modelling

One exonic variant in *TGM6* (GRCh37/hg19, NC_000020.10:g.2398091T>G, NM_198994.2:c.1550T>G) (Figure 3) can be validated, which was identified by both of the targeted NGS and WES as the sole coding variant within the linkage region. The exonic variants in *CENPB* and *SIRPA* can not be validated by sanger sequencing and may be NGS false positives. In addition, we observed that the *TGM6* mutation was present in 2 patients and 6 members at potential preleukaemic phase, but not in 13 unaffected individuals, and it was also absent in 530 ethnically matched healthy controls. These data suggest that the variant cosegregated with the disease in this family. The mutation was located in a highly conserved position throughout vertebrates including amphibians (Figure 4), and was predicted to be functionally damaging using SIFT and PolyPhen software. Sequence alignments by Promals3D revealed that *TGM6* was highly similar to *TGM2* and *TGM3*, which therefore can be used as templates for *TGM6* modelling. We observed the L517W mutation located in the first β -barrel domain of *TGM6* molecular structure (Supplementary Figure 2), which was important for the conformational transition from an inactive compact form to an active, extended ellipsoid structure¹⁶ that exposes the TGM catalytic core. In addition, this

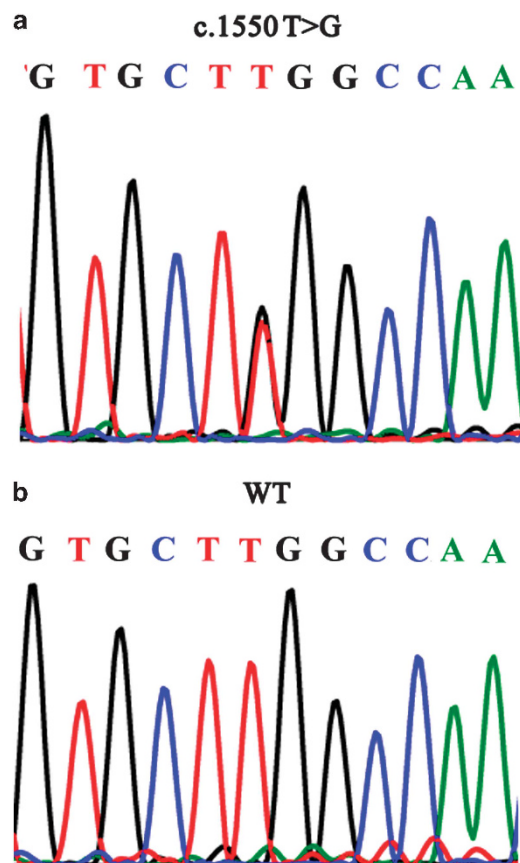


Figure 3 Validation of the *TGM6* mutation. (a) Sanger sequencing of the patient III-15 confirmed the presence of the *TGM6* mutation. (b) Sanger sequencing of an unaffected family member suggested a wide-type *TGM6*.

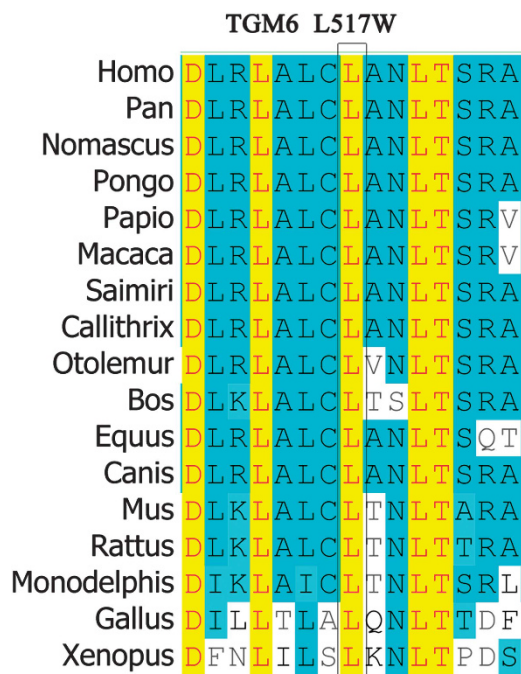


Figure 4 Alignment of *TGM6* amino acid sequences in 17 vertebrates, which suggest the affected amino-acid L517 was located in a highly conserved position. The left column represents the species and the right shows the amino-acid sequence in the corresponding species; amino acids that identical to those in *Homo sapiens* are highlighted in yellow; conservative to homo sapiens are highlighted in blue; weakly similar or non-similar are not highlighted.

mutation was near the GDP/GTP-binding pocket, which may interfere with the allosteric regulation of GDP/GTP and affect *TGM6* activation.

DISCUSSION

Familial aggregation of AML is exceedingly rare. A large family with 13 cases in four generations was previously documented, but contact with this family was lost in 1980.¹⁷ The family in our study constituted the second largest reported pedigree worldwide, with 11 cases in four generations who transmitted AML in an autosomal-dominant manner. The age at AML onset decreased with each passing generation, which is consistent with previous reports¹⁸ and suggests that the genetic factor was the primary pathogenic factor in this family. We excluded common constitutional cytogenetic abnormalities, such as -7 , $+8$, $-5q$, $+21$, which frequently occurred in MDS/AML. Besides, no known causative mutations were found in patients of this family. These data strongly suggest that novel genetic variants may be responsible for the disease in this family.

Owing to the high penetrance of AML in this family, we hypothesised that the disease may result from single-gene inborn errors. The combination of linkage analysis with recently developed NGS technology have greatly accelerated the discovery of novel susceptible genes in rare Mendelian disorders.¹⁹ Using this combined strategy, we identified a previously unreported candidate region linked to 20p13 in our family, with a maximum multipoint HLOD score of 3.56 ($P=0.00005$). Subsequent targeted NGS of the linkage interval revealed a missense mutation in *TGM6* (L517W) that cosegregated with the phenotype in this family, and was absent in 530 healthy controls and the dbSNP138 and 1000 Genome Project

databases. These suggest that *TGM6* may be a candidate gene for familial AML.

However, we can not simply exclude other exonic variants within the linkage region in terms of the coverage of all exons in the targeted region. So we performed additional WES to further explore the coding variants within the candidate region. Interestingly, WES identified the same variant in *TGM6* as the sole coding variant within the linkage region, which again support its candidacy in familial AML.

This study is the first to implicate a *TGM6* mutation in leukaemia, however, reports on the role of the *TGM* family in tumours were not uncommon. Numerous studies have demonstrated that *TGM2* expression is downregulated in primary tumours. The upregulation of *TGM2* and intratumour *TGM2* injections inhibit tumour growth in mice.¹⁶ The reduction or loss of *TGM3* expression is common in oesophageal, laryngeal, oral, head and neck squamous cell carcinomas, partially due to its important role in the regulation of stratified squamous epithelia differentiation.²⁰ Research on *FXIII-A-/-* mice suggests that factor XIII transglutaminase supports haematogenous tumour cell metastasis.²¹ These results suggest that *TGM* family members are widely involved in multistep tumour development.

TGM6 is a newly identified member of the *TGM* family that exhibits a high structural similarity to *TGM2* and *TGM3*, which post-translationally modify proteins by catalysing a Ca^{2+} -dependent transferase reaction, and are allosterically regulated by Ca^{2+} and GDP/GTP.¹⁶ Several studies demonstrate that retinoic acid (RA)-induced differentiation of myeloid leukaemia cell lines, such as HL60, HEL and THP1, is frequently accompanied by a marked increase in transglutaminase activity, whereas transglutaminase is nearly undetectable before RA treatment in these cell lines. The knockdown of transglutaminase expression using specific siRNA or a transglutaminase inhibitor abrogates the effect of RA.^{22–25} Transglutaminase activity is also greatly increased during induced differentiation in various types of tumours.²⁶ These data suggest that transglutaminase activity has a significant role in cell differentiation. Interestingly, the three genes, *RUNX1*, *CEBPA* and *GATA2*, which have been identified in AML-predisposed families were transcription factors that regulate myeloid differentiation. The patients in the family we studied were diagnosed with different AML subtypes, which suggest that the pathogenic genes may be implicated in the regulation of early myeloid differentiation. Combining these findings, we propose that *TGM6* may participate in leukaemogenesis because of the role of transglutaminase activity in cell differentiation.

The variation in *TGM6* was at a highly conserved position, and predicted as deleterious using SIFT and Polyphen software. The affected amino acid was located in the first β -barrel domain and close to the GDP/GTP-binding pocket, which is important for the conformational transition from an inactive compact form to an active extended ellipsoid structure.¹⁶ The amino-acid substitution from leucine to tryptophan may interfere with the conformation change and decrease or completely eliminate transglutaminase activity of *TGM6* and produce a haploinsufficiency of *TGM6* function. Given the apparent pattern of autosomal-dominant hereditary and the absent of copy number variants in patients of our family (unpublished data), the haploinsufficiency of *TGM6* function may be a possible mechanism underlying the leukaemia in this family.

Recently, *TGM6* mutations have been identified in two autosomal-dominant spinocerebellar ataxia (SCA) families using a combined strategy of exome sequencing and linkage analysis.¹⁵ In addition, we also found a deleterious *TGM6* mutation in an autosomal-dominant AML family, which proved to be cosegregated with the disease

phenotype in the family. These facts suggested that *TGM6* may be a pathogenic factor for both familial AML and SCA; and there have been numerous reports of mutations of one gene causing a disease with a wide range of symptoms including ataxia and blood malignancies, such as *ATM* mutations causing ataxia-telangiectasia,²⁷ and *TERC/TERT* mutations causing dyskeratosis congenital.²⁸ In fact, two members of *TGM* family, *TGM2* and *FXIII-A*,¹⁶ have been recognised as pleiotropy genes and can affect multiple traits. So we suggest that *TGM6* may be another pleiotropy gene, and contribute to both ataxia and leukaemia. However, no ataxia patients have been found in our family and no leukaemia patients found in the SCA families mentioned above. Owing to limited literatures, more exhausted efforts are needed to determine the *TGM6* mutation distribution in different disorders and to explore the potential modified genes that synergise with *TGM6* in specific disease.

In addition, we found that WES is a cost-effective solution to detect exonic variants within the candidate region because it had a similar performance like targeted NGS. They both detected 52–58 exonic variants per sample in our study. However, WES would skip a large number of intronic, regulatory and UTR variants that may also contribute to the disease aetiology. So we suggest that targeted NGS would be a preferable solution if specific candidate regions have been identified. However, WES would be better if such regions were absent and enough samples were available to filter a huge amount of SNPs. In this study, we combined two approaches of NGS and detected a large number of intronic, intergenic and UTR variants. Considering most of causative variants occurred in exons, we primarily focus on exonic variants and finally detected a coding variant in *TGM6* after a stringent process of filtration and validation.

In conclusion, we identified a previously unreported linkage region on 20p13. Subsequent WES and targeted NGS identified a missense mutation in *TGM6* as the sole exonic variant within this region. Combining bioinformatic analysis and literature reports, we suggest *TGM6* as a novel candidate gene that may be associated with familial AML, and its discovery will promote the understanding of the role of transglutaminases in leukaemogenesis. However, further studies are required to expand on this theoretical foundation. Our study again proves the efficiency of the combined strategy of linkage analysis and NGS technology in identifying candidate genes in rare Mendelian disorders, and targeted NGS and WES have advantages of their own in specific conditions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all of the family members and volunteers for their participation in this study. We further thank Dr Jiucun Wang and Dr Qiang Huang (School of Life Sciences, Fudan University) for their technology support. Thanks are also due to Dr Zuwei Qian of the Affymetrix Company for offering 8 sets of the 500K SNP array. This work was supported by the National Natural Science Foundation of China (81270609, 30770909) and the Major Science and Technology Project of Fujian Province (2003F003, 2012Y4012) and Fujian Medical University (09ZD008).

- 1 Song WJ, Sullivan MG, Legare RD *et al*: Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* 1999; **23**: 166–175.
- 2 Hahn CN, Chong CE, Carmichael CL *et al*: Heritable *GATA2* mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 2011; **43**: 1012–1017.
- 3 Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J: Mutation of *CEBPA* in familial acute myeloid leukemia. *N Engl J Med* 2004; **351**: 2403–2407.
- 4 Kirwan M, Vulliamy T, Marrone A *et al*: Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia. *Hum Mutat* 2009; **30**: 1567–1573.
- 5 Holme H, Hossain U, Kirwan M, Walne A, Vulliamy T, Dokal I: Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br J Haematol* 2012; **158**: 242–248.
- 6 Quintana-Bustamante O, Lan-Lan Smith S, Griessinger E *et al*: Overexpression of wild-type or mutants forms of *CEBPA* alter normal human hematopoiesis. *Leukemia* 2012; **26**: 1537–1546.
- 7 He LZ, Lu LH, Chen ZZ: Genetic mechanism of leukemia predisposition in a family with 7 cases of acute myeloid leukemia. *Cancer Genet Cytogenet* 1994; **76**: 65–69.
- 8 Matsuzaki H, Loi H, Dong S *et al*: Parallel genotyping of over 10 000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* 2004; **14**: 414–425.
- 9 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 10 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 11 Rehman AU, Morell RJ, Belyantseva IA *et al*: Targeted capture and next-generation sequencing identifies *C9orf75*, encoding taperin, as the mutated gene in nonsyndromic deafness *DFNB79*. *Am J Hum Genet* 2010; **86**: 378–388.
- 12 Hedges DJ, Burges D, Powell E *et al*: Exome sequencing of a multigenerational human pedigree. *PLoS One* 2009; **4**: e8232.
- 13 Ng PC, Henikoff S: Predicting deleterious amino acid substitutions. *Genome Res* 2001; **11**: 863–874.
- 14 Sunyaev S, Ramensky V, Koch I, Lathe 3rd W, Kondrashov AS, Bork P: Prediction of deleterious human alleles. *Hum Mol Genet* 2001; **10**: 591–597.
- 15 Wang JL, Yang X, Xia K *et al*: *TGM6* identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* 2010; **133**: 3510–3518.
- 16 Iismaa SE, Mearns BM, Lorand L, Graham RM: Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. *Physiol Rev* 2009; **89**: 991–1023.
- 17 Carmichael CL, Wilkins EJ, Bengtsson H *et al*: Poor prognosis in familial acute myeloid leukaemia with combined biallelic *CEBPA* mutations and downstream events affecting the *ATM*, *FLT3* and *CDX2* genes. *Br J Haematol* 2010; **150**: 382–385.
- 18 Horwitz M, Goode EL, Jarvik GP: Anticipation in familial leukemia. *Am J Hum Genet* 1996; **59**: 990–998.
- 19 Grossmann V, Kohlmann A, Klein HU *et al*: Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. *Leukemia* 2011; **25**: 671–680.
- 20 Liu W, Yu ZC, Cao WF, Ding F, Liu ZH: Functional studies of a novel oncogene *TGM3* in human esophageal squamous cell carcinoma. *World J Gastroenterol* 2006; **12**: 3929–3932.
- 21 Palumbo JS, Barney KA, Blevins EA *et al*: Factor XIII transglutaminase supports hematogenous tumor cell metastasis through a mechanism dependent on natural killer cell function. *J Thromb Haemost* 2008; **6**: 812–819.
- 22 Davies PJ, Murtaugh MP, Moore Jr WT, Johnson GS, Lucas D: Retinoic acid-induced expression of tissue transglutaminase in human promyelocytic leukemia (HL-60) cells. *J Biol Chem* 1985; **260**: 5166–5174.
- 23 Suedhoff T, Birckbichler PJ, Lee KN, Conway E, Patterson Jr MK: Differential expression of transglutaminase in human erythroleukemia cells in response to retinoic acid. *Cancer Res* 1990; **50**: 7830–7834.
- 24 Mehta K, Lopez-Berestein G: Expression of tissue transglutaminase in cultured monocytic leukemia (THP-1) cells during differentiation. *Cancer Res* 1986; **46**: 1388–1394.
- 25 Singh US, Kunar MT, Kao YL, Baker KM: Role of transglutaminase II in retinoic acid-induced activation of RhoA-associated kinase-2. *EMBO J* 2001; **20**: 2413–2423.
- 26 Lentini A, Provenzano B, Tabolacci C, Beninati S: Protein-polyamine conjugates by transglutaminase 2 as potential markers for antineoplastic screening of natural compounds. *Amino Acids* 2009; **36**: 701–708.
- 27 Savitsky K, Bar-Shira A, Gilad S *et al*: A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 1995; **268**: 1749–1753.
- 28 Tsangaris E, Adams SL, Yoon G *et al*: Ataxia and pancytopenia caused by a mutation in *TINF2*. *Hum Genet* 2008; **124**: 507–513.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)