Polymorphism in TGFB1 is associated with worse non-relapse mortality and overall survival after stem cell transplantation with unrelated donors

Esteban Arrieta-Bolaños,[1,2,3] Neema P. Mayor,[1,2] Steven G.E. Marsh,[1,2] J. Alejandro Madrigal,[1,2] Jane F. Apperley,[4] Keiren Kirkland,[5] Stephen Mackinnon,[6] David I. Marks,[7] Grant McQuaker,[8] Julia Perry,[5] Michael N. Potter,[9] Nigel H. Russell,[10] Kirsty Thomson,[11] and Bronwen E. Shaw[1,2]

[1]Anthony Nolan Research Institute, London, UK; [2]Cancer Institute, University College London, UK; [3]Centro de Investigaciones en Hematología y Trastornos Afines (CIHATA), Universidad de Costa Rica, San José, Costa Rica; [4]Haematology Department, Hammersmith Hospital, London, UK; [5]BSBMT Data Registry, Guy's Hospital, London, UK; [6]Department of Haematology, University College London, Royal Free Campus, UK; [7]Adult BMT Unit, University Hospitals Bristol NHS Trust, Bristol, UK; [8]Bone Marrow Transplant Unit, Beatson West of Scotland Cancer Centre, Glasgow, UK; [9]Section of Haemato-oncology, Royal Marsden Hospital, London, UK; [10]Centre for Clinical Haematology, Nottingham University Hospital, and Academic Haematology, Nottingham University Hospitals, UK; and [11]Department of Haematology, University College Hospital, London, UK

 doi:10.3324/haematol.2015.134999

Received: August 10, 2015.

Accepted: November 20, 2015.

Pre-published: November 26, 2015.

Correspondence: a.madrigal@ucl.ac.uk

Supplementary methods

HSCT patient and donor samples were subjected to typing using the strategy described in (6). In brief, genomic DNA was used to separately amplify 4 sub-regions of *TGFB1*'s regulatory region and exon 1. These 4 amplicons were then purified in order to be used as templates for bidirectional Sanger sequencing using the M13 tags attached to the amplification primers. The sequences were then analyzed and the genotypes for each of the 18 basic polymorphic positions were recorded and input in a database. Additionally, the rest of the sequence was scanned for the presence of potentially novel polymorphisms outside the basic 18 sites. Based on the genotype at these basic polymorphic positions and the alleles previously described in isolation by Shah and collaborators (4,5, **Supplementary table 1**), the allele combination for each sample was inferred, and an allelic genotype was assigned to each patient or donor. For example, in a sample with these genotypes:

-2410	-2389	-1985	-1638	-1347	-1287	-1169	-1154	-827	-778	-768	-469	-387	-229	-14	+29	+74	+91
AG	AGG/AGG	CC	GG	CC	GG	TT/TT	CC	GG	GG	-/C	CC	CC	CC	GG	TC	GC	TT

the presence of allele p017 would be discarded by position +91, then allele p005 would be discarded by position -14, allele p015 by position -229, and so on. Heterozygous positions would then be probed with combinations of the alleles not discarded by the homozygous positions in order to determine the final genotype (in this example p003/p014).

In cases where theoretical ambiguities existed, the phase of the relevant polymorphic positions was defined by allele-specific amplification strategies using different primer combinations as described in (6). For example, when the genotype included a heterozygous result for the -2389dupAGG, then the -2389dupAGG-specific primers were used instead of the generic forward primer in combination with any of the reverse primers. The choice of reverse primer depended on the location of the relevant polymorphisms in order to create a -2389AGG-specific amplicon that would set their phase. Likewise, the +29 T>C-specific primers were also used to amplify either the +29T-bearing or the +29C-bearing allele and set the phase of other relevant polymorphisms that would define the allelic combination. This strategy was also used in the validation stages for the typing protocol using samples from healthy volunteer donors and whole-region amplification, prior to the typing of the clinical samples.

Ten percent of the patient and donor samples were selected to confirm their typing results. For this, a new amplification from the stock DNA was carried out and sequencing of Region IV was repeated. All of these samples had results that were consistent with the original ones.

Overall, the 1,024 samples typed showed polymorphism for 9 of the 18 previously known variable positions. These were -2410A>G, -2389dupAGG, -1638G>A, -1347C>T, -1169delTT, -778G>A, -768insC, +29T>C and +74G>C. The results for the variant and genotype frequencies observed for each of these positions are shown in **Supplementary table 2**. Two of these positions (-1169delTT and -778G>A) showed extremely rare variation, leaving the other 7 positions as the most frequently polymorphic.

Deviation from Hardy-Weinberg equilibrium (HWE) was tested for 5 of the SNPs (**Supplementary table 3**). The two most infrequent polymorphisms were excluded because of their rarity. From polymorphisms that show the same observed genotypic frequencies (i.e. -2389dupAGG and -1347C>T, and -768insC and +74G>C) only one was selected. *TGFB1* +29T>C showed a significant deviation from HWE (p=0.03) caused by an excess of heterozygotes. Deviation from HWE was also tested for the 4 major *TGFB1* regulatory region and exon 1 alleles (**Supplementary table 4**). The genotypic distributions were found not to be significantly deviated from expected frequencies (p>0.25).

These polymorphic positions allowed for the definition of the allelic genotypes for each patient-donor pair. Among the 17 previously defined regulatory region and exon 1 alleles, only 6 were seen in the cohort: p001, p003, p006, p009, p013, and p014. Moreover, only 4 were the predominant ones: p003 (53.71%), p001 (29.35%), p014 (8.25%), and p006 (8.11%). p019 and p013 were only seen in 1 and 2 samples, respectively. The genotype frequencies observed in the cohort are presented in **Supplementary table 5**. Twelve genotypes formed by the known *TGFB1* regulatory region and exons 1 were observed. Both the allele and the genotype frequencies did not differ significantly between patients and donors (Z test; p>0.05).

Supplementary table 1 ***TGFB1*** **promoter alleles based on 18 polymorphic positions within its upstream regulatory region and exon 1 as reported by Shah *et al*. (2006, 2009)**

Allele[b]	Position[a]																	
	-2410	-2389	-1985	-1638	-1347	-1287	-1169	-1154	-827	-778	-768	-469	-387	-229	-14	+29	+74	+91
p001	A	-	C	G	T	G	TT	C	G	G	-	C	C	C	G	C	G	T
p002	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p003	G	AGG	-	-	C	-	-	-	-	-	-	-	-	-	-	T	-	-
p004	G	AGG	G	-	C	-	-	-	-	-	-	-	-	-	-	T	-	-
p005	-	AGG	-	-	C	A	-	T	C	-	-	-	T	-	A	-	-	-
p006	G	AGG	-	A	C	-	-	-	-	-	-	-	-	-	-	T	-	-
p007	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p008	G	AGG	-	-	C	-	-	-	-	-	-	A	-	-	-	T	-	-
p009	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-
p010	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-
p011	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	T	-	-
p012	-	-	-	-	-	-	del	-	-	-	-	-	-	-	-	-	-	-
p013	G	AGG	-	-	C	-	del	-	-	-	-	-	-	-	-	T	-	-
p014	-	AGG	-	-	C	-	-	-	-	-	C	-	-	-	-	-	C	-
p015	G	AGG	-	-	-	-	-	-	-	-	-	-	-	G	-	T	-	-
p016	G	AGG	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-
p017	G	-	-	-	C	-	-	-	-	-	-	-	-	-	-	T	-	G

a. Nucleotide position relative to the major translation start site (+1). Only positions that are polymorphic are included.
b. Promoter allele defined by the sequence of the regulatory region and exon 1 as described. Sequences are compared to allele p001 (GenBank accession no. AY871232).
del, deletion of nucleotides. Modified from Shah *et al*. (2006 and 2009)[4,5].

Supplementary table 2 Variant and genotype frequencies for known *TGFB1* regulatory region and exon 1 polymorphisms found in the UD-HSCT patient-donor cohort.

Polymorphic position	Position in chromosome 19[a]	Variants and genotypes	Whole cohort		Patients		Donors		Z	p value
			count	frequency	count	frequency	count	frequency		
-2410G>A	41355454	G	1269	**0.6196**	632	**0.6184**	637	**0.6209**	0.1	0.91
(rs4803457)		A	779	**0.3804**	390	**0.3816**	389	**0.3791**	0.1	0.91
		GG	378	**0.3691**	186	**0.3640**	192	**0.3743**	0.3	0.73
		AG	513	**0.5010**	260	**0.5088**	253	**0.4932**	0.5	0.62
		AA	133	**0.1299**	65	**0.1272**	68	**0.1326**	0.3	0.80
-2389dupAGG	41355432	AGG	1439	**0.7026**	712	**0.6967**	727	**0.7086**	0.6	0.56
(rs11466313)		…	609	**0.2974**	310	**0.3033**	299	**0.2914**	0.6	0.56
		AGG/AGG	499	**0.4873**	246	**0.4814**	253	**0.4932**	0.4	0.71
		AGG/…	441	**0.4307**	220	**0.4305**	221	**0.4308**	0.0	0.99
		…/…	84	**0.0820**	45	**0.0881**	39	**0.0760**	0.7	0.49
-1638G>A	41354682	G	1882	**0.9189**	945	**0.9247**	937	**0.9133**	0.9	0.34
(rs1800468)		A	166	**0.0811**	77	**0.0753**	89	**0.0867**	0.9	0.34
		GG	863	**0.8428**	438	**0.8571**	425	**0.8285**	1.3	0.21
		GA	156	**0.1523**	69	**0.1350**	87	**0.1696**	1.5	0.12
		AA	5	**0.0049**	4	**0.0078**	1	**0.0019**	N/A	N/A
-1347C>T	41354391	C	1439	**0.7026**	712	**0.6967**	727	**0.7086**	0.6	0.56
(rs1800469)		T	609	**0.2974**	310	**0.3033**	299	**0.2914**	0.6	0.56
		CC	499	**0.4873**	246	**0.4814**	253	**0.4932**	0.4	0.71
		CT	441	**0.4307**	220	**0.4305**	221	**0.4308**	0.0	0.99
		TT	84	**0.0820**	45	**0.0881**	39	**0.0760**	0.7	0.49
-1169delTT	41354213	TT	2046	**0.9990**	1020	**0.9980**	1026	**1.0000**	N/A	N/A
(rs56368056)		..	2	**0.0010**	2	**0.0020**	0	**0.0000**	N/A	N/A
		TT/TT	1022	**0.9980**	509	**0.9961**	513	**1.0000**	N/A	N/A
		TT/..	2	**0.0020**	2	**0.0039**	0	**0.0000**	N/A	N/A
		../..	0	**0.0000**	0	**0.0000**	0	**0.0000**	N/A	N/A
-778G>A	41353822	G	2047	**0.9995**	1022	**1.0000**	1025	**0.9990**	N/A	N/A
(rs36185305)		A	1	**0.0005**	0	**0.0000**	1	**0.0010**	N/A	N/A
		GG	1023	**0.9990**	511	**1.0000**	512	**0.9981**	N/A	N/A
		GA	1	**0.0010**	0	**0.0000**	1	**0.0019**	N/A	N/A
		AA	0	**0.0000**	0	**0.0000**	0	**0.0000**	N/A	N/A
-768insC	41353812	.	1879	**0.9175**	943	**0.9227**	936	**0.9123**	0.9	0.39
(rs1800999)		C	169	**0.0825**	79	**0.0773**	90	**0.0877**	0.9	0.39
		./.	862	**0.8418**	436	**0.8532**	426	**0.8304**	1.0	0.32
		./C	155	**0.1514**	71	**0.1389**	84	**0.1637**	1.1	0.27
		C/C	7	**0.0068**	4	**0.0078**	3	**0.0058**	N/A	N/A
+29T>C	41353016	T	1275	**0.6226**	635	**0.6213**	640	**0.6238**	0.1	0.91
(rs1800470)		C	773	**0.3774**	387	**0.3787**	386	**0.3762**	0.1	0.91
		TT	380	**0.3711**	187	**0.3659**	193	**0.3762**	0.3	0.73
		TC	515	**0.5029**	261	**0.5108**	254	**0.4951**	0.5	0.62
		CC	129	**0.1260**	63	**0.1233**	66	**0.1287**	0.3	0.79
+74G>C	41352971	G	1879	**0.9175**	943	**0.9227**	936	**0.9123**	0.9	0.39
(rs1800471)		C	169	**0.0825**	79	**0.0773**	90	**0.0877**	0.9	0.39
		GG	862	**0.8418**	436	**0.8532**	426	**0.8304**	1.0	0.32
		GC	155	**0.1514**	71	**0.1389**	84	**0.1637**	1.1	0.27
		CC	7	**0.0068**	4	**0.0078**	3	**0.0058**	N/A	N/A

Only polymorphic positions are shown. del, deletion; dup, duplication; ins, insertion. A dot (.) indicates a deletion or the absence of insertion. Z test and associated probability (p) for the comparison of frequencies between patient and donor subsets are shown. N/A, non-applicable. [a]Based on assembly GRCh38, build 106.

Supplementary table 3 **Analysis of Hardy-Weinberg equilibrium for selected *TGFB1* regulatory region and exon 1 SNPs in the patient-donor cohort.**

Polymorphic position	Genotype	Observed numbers	Expected numbers	p value (Fisher's exact test)
-2410G>A	GG	378	393.15	0.05
	AG	513	482.69	
	AA	133	148.15	
-1347C>T	CC	499	505.55	0.37
	CT	441	427.91	
	TT	84	90.55	
+29T>C	TT	380	396.88	0.03
	TC	515	481.24	
	CC	129	145.88	
+74G>C	GG	862	861.97	1.00
	GC	155	155.05	
	CC	7	6.97	

Supplementary table 4 **Analysis of Hardy-Weinberg equilibrium for *TGFB1* regulatory region and exon 1 alleles in the patient-donor cohort.**

Genotype	Observed numbers	Expected numbers	p value (Chi^2)
p001/p001	81	88.05	>0.25
p001/p003	346	322.69	
p001/p006	49	48.96	
p001/p014	40	49.26	
p003/p003	280	295.66	
p003/p006	91	89.72	
p003/p014	97	90.27	
p006/p006	5	6.81	
p006/p014	16	13.70	
p014/p014	7	6.89	

Supplementary table 5 *TGFB1* regulatory region and exon 1 genotype frequencies found in the UD-HSCT patient-donor cohort.

Genotype	Whole cohort		Patients[1]		Donors		Z	p value
	count	frequency	count	frequency	count	frequency		
p001/p001	81	**0.0791**	43	**0.0841**	38	**0.0741**	0.6	0.55
p001/p003	346	**0.3379**	182	**0.3562**	164	**0.3197**	1.2	0.22
p001/p006	49	**0.0479**	21	**0.0411**	28	**0.0546**	1.0	0.31
p001/p014	40	**0.0391**	16	**0.0313**	24	**0.0468**	1.3	0.20
p003/p003	280	**0.2734**	138	**0.2701**	142	**0.2768**	0.2	0.81
p003/p006	91	**0.0889**	43	**0.0841**	48	**0.0936**	0.5	0.59
p003/p014	97	**0.0947**	49	**0.0959**	48	**0.0936**	0.1	0.90
p006/p006	5	**0.0049**	4	**0.0078**	1	**0.0019**	N/A	N/A
p006/p014	16	**0.0156**	5	**0.0098**	11	**0.0214**	1.5	0.13
p014/p014	7	**0.0068**	4	**0.0078**	3	**0.0058**	N/A	N/A
p003/p009	1	**0.0010**	0	**0.0000**	1	**0.0019**	N/A	N/A
p003/p013	2	**0.0020**	2	**0.0039**	0	**0.0000**	N/A	N/A
Other[2]	9	**0.0088**	4	**0.0078**	5	**0.0097**	N/A	N/A
Total	1024	**1.0000**	511	**1.0000**	513	**1.0000**		

[1] 18 pairs were excluded from analyses because of lack of clinical data. Patients excluded bore *TGFB1* p001/p003 (n=7), p003/p003 (n=6), p001/p001 (n=2), p006/p014 (n=2), and p014/p014 (n=1) genotypes. [2] Other genotypes formed by one previously described allele and one of three novel alleles found in the cohort. Z test and associated probability (p) for the comparison of frequencies between patient and donor subsets are shown. N/A, non-applicable.