SUPPLEMENTARY MATERIALS

Figure S1: Sanger sequencing validation of 9 non-*TGFBR1* variants identified in Scottish MSSE families with common ancestry

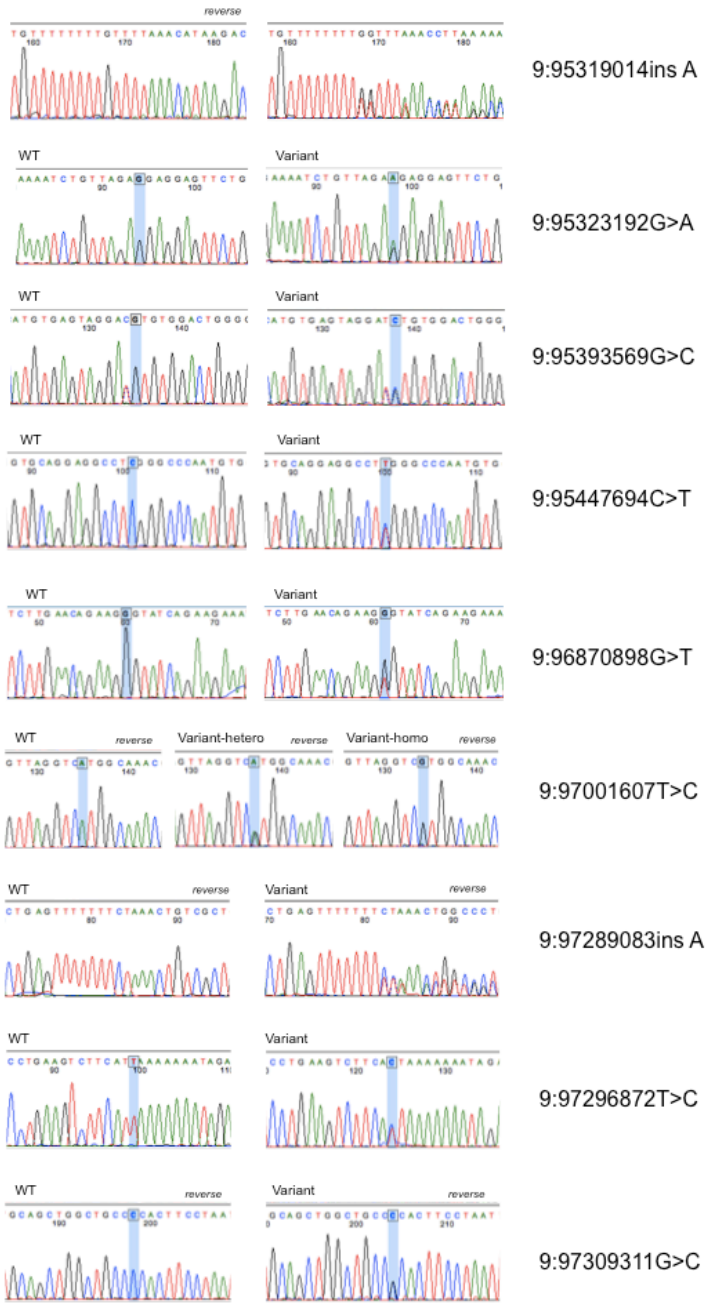

Figure S2: Sanger sequencing validation of 8 variants identified in Scottish family 17

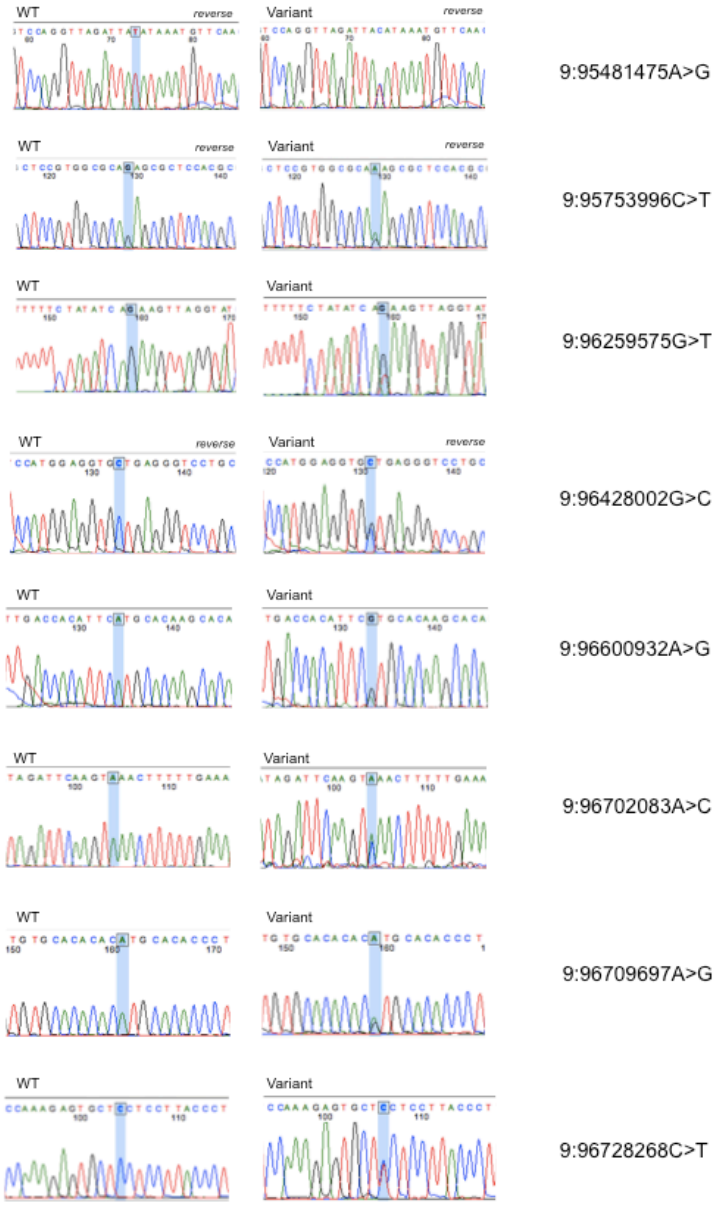

Figure S3: Sanger sequencing validation of 7 variants identified in Scottish family 15

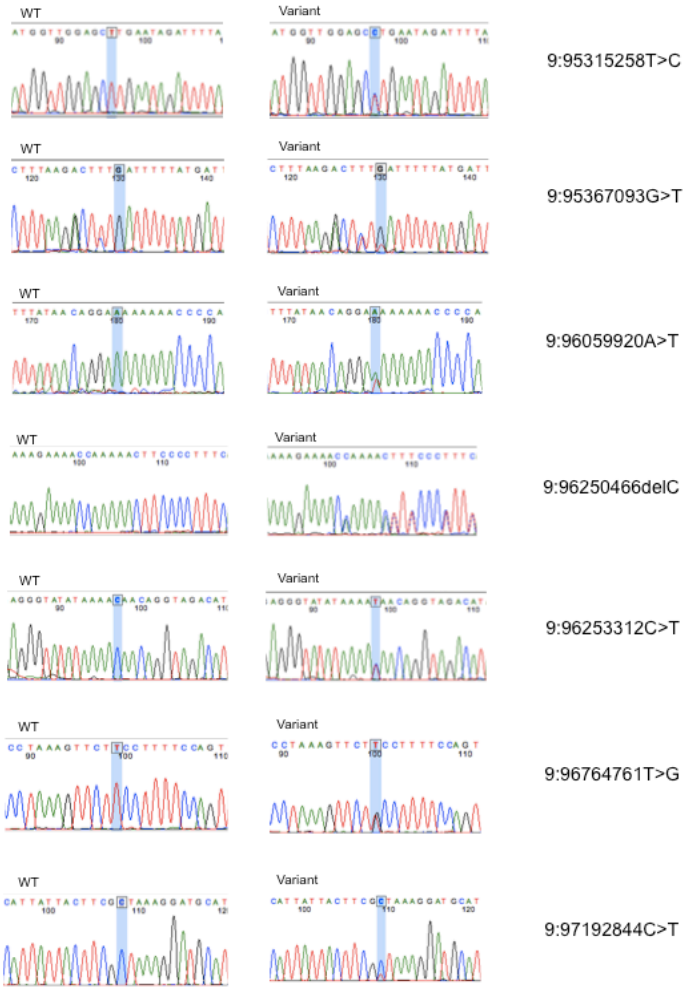

Table S1: Summary of 13 MSSE families available in this study

Family no.[1]	*TGFBR1* mutation[1]	Shared 9q22.3 haplotype[1]	9 rare variants at non-*TGFBR1* locus	Sequencing (n)[2]
1	p.C41Y	No	No	Sanger[3] (1)
2	p.N45S	Yes, but diverged	Yes	Sanger (1)
3	c.134-135insA	No	No	Illumina[4] (1)
4	p.G52R	Yes	Yes	Illumina (3), Sanger (4)
5	p.G52R	Yes	Yes	Illumina (1), Sanger (3)
6	p.G52R	Yes	Yes	Illumina (1)
7	p.G52R	Yes	Yes	Sanger (1)
9	p.G52R	Yes	Yes	Illumina (1)
13	p.W242X	No	No	Illumina (1), Sanger (1)
14	c.806-2A>C	No	No	Illumina (1)
15	c.980delC	No	No	Illumina (2)
17	c.1059_1062delACTGinsCAATAA	No	No	Illumina (2)
18	p.R414X	Yes, but diverged	Yes	Illumina (1), Sanger (2)

Abbreviations: no.,number

[1]Goudie et al., Nature Genetics, 2011

[2]Indicates the number of affected individuals sequenced

[3]Sequenced by Sanger dideoxy sequencing

[4]First, sequenced by Illumina Genome Analyzer II and confirmed by Sanger dideoxy sequencing

Table S2: Seven MSSE-associated non-*TGFBR1* rare variants identified in Scottish family 15

Chr Positions (NCBI36)	Chr Positions (GRCh37)	Gene	Region	W	V	Family 15 MA-BM	Family 15 MA-1803	Normal (n=162)[2]
95315258	96275437	*FAM120A*	intronic	T	C[1]	T/C	T/C	1/162
95367093	96327272	*FAM120A*	3'UTR	G	T	G/T	G/T	0/160
96059920	97020099	*LOC158257*	ncRNA	A	T	A/T	A/T	0/162
96250466	97210646	*HIATL1*	intronic	C	del C[1]	C/-	C/-	5/160
96253312	97213491	*HIATL1*	intronic	C	T	C/T	C/T	0/162
96764761	97724940	*C9orf3*	intronic	T	G[1]	T/G	T/G	9/162
97192844	98153023	NA	intergenic	C	T[1]	C/T	C/T	1/162

Abbreviations: Chr, chromosome; UTR, untranslated region; 1000G; 1000 Genomes, W, wild-type; V, variant

[1]Variants are reported in 1000 Genome (http://www.1000genomes.org).

[2]162 unrelated healthy normal controls from Scottish population

Table S3: Eight MSSE-associated non-*TGFBR1* rare variants identified in Scottish family 17

Chr Positions		Gene	Region	W	V	Family 17				Normal
(NCBI36)	(GRCh37)					LE-3	LE-4	LE-5	LE-94	(N=162)[2]
95481475	96441654	*PHF2*	3'UTR	A	G	A/A	A/G	A/G	A/G	1/150
95753996	96714175	*BARX1*	3'UTR	C	T	C/C	C/T	C/T	C/T	1/160
96259575	97219754	*HIATL1*	intronic	G	T	G/T	G/T	G/T	G/T	1/162
96428002	97388181	*FBP1*	intronic	G	C	G/C	G/C	G/C	G/C	1/154
96600932	97561111	*C9orf3*	intronic	A	G[1]	A/G	A/G	A/G	A/G	5/148
96702083	97662262	*C9orf3*	intronic	A	C	A/C	A/C	A/C	A/C	1/161
96709697	97669876	*C9orf3*	intronic	A	G[1]	A/G	A/G	A/G	A/G	4/160
96728268	97688447	*C9orf3*	intronic	C	T[1]	C/T	C/T	C/T	C/T	4/160

Abbreviations: Chr, chromosome; UTR, untranslated region; 1000G; 1000 Genomes, W, wild-type; V, variant

[1]Variants are reported in 1000 Genome (http://www.1000genomes.org).

[2]162 unrelated healthy normal controls from Scottish population

One healthy unrelated individual from Scotland has all 8 variants identified in Family 17.

Table S4: Primer sequences used for sequencing 9 rare variants identified in Scottish MSSE families with common ancestry

Primer name	Sequence (5' to 3')
95319014F	CCCAAGCAAAACCCTAGAAA
95319014R	AGGAAAAGATCACCTGCTTCA
95323192F	TTGGATGAGTTGTTCAGTGTCTTT
95323192R	TTCCAACAGGTTCACCTTCA
95393569F	TTTGGAAAGGAAGATGTAATGAAA
95393569R	CCAGGGTCTGGAGTCTTCCT
95447694F	GCTGCTCTATCGTCCCATTT
95447694R	TGCACAGTGGTGGGTGTC
96870898F	GCCTGTGTCTTTCCTAAAAGTTG
96870898R	ACGCACACTGAGGCCTAAAA
97001607F	CACCCCTCCATTCAACAAAT
97001607R	AAGCCTTCCCAGGATTGACT
97289083F	CGCTGTTCCCAGTCCTTATT
97289083R	CTCCCCCACTCCACAGTCTA
97296872F	GGCTGGATAATGGCAAGAAG
97296872R	AGAACCCAGGAACCAACACA
97309311F	ATGTGTGGAATCCAGGGAAG
97309311R	AGCGCTCTTTGGAACAACAT

Table S5: Genes located in the SRH locus

Gene symbol	Chr	start position	stop position	size (bp)	Description
FAM120A	9	95,253,994	95,368,218	114,224	family with sequence similarity 120A
PHF2	9	95,378,730	95,481,690	102,960	PHD finger protein 2
BARX1	9	95,753,730	95,757,429	3,699	BARX homeobox 1
PTPDC1	9	95,832,897	95,911,957	79,060	protein tyrosine phosphatase domain containing 1
CYCSP24	9	95,840,133	95,840,448	315	cytochrome c, somatic pseudogene 24
MIRNLET7A1	9	95,978,059	95,978,138	79	microRNA let-7a-1
MIRNLET7F1	9	95,978,449	95,978,535	86	microRNA let-7f-1
LOC158257	9	95,978,561	96,061,936	83,375	hypothetical protein LOC158257
MIRNLET7D	9	95,980,936	95,981,022	86	microRNA let-7d
ZNF169	9	96,061,399	96,105,112	43,713	zinc finger protein 169
FAM22F	9	96,120,299	96,130,747	10,448	family with sequence similarity 22, member F
LOC100132077	9	96,141,005	96,162,011	21,006	hypothetical protein LOC100132077
LOC728026	9	96,148,562	96,150,667	2,105	hypothetical protein LOC728026
HIATL1	9	96,176,654	96,263,023	86,369	hippocampus abundant transcript-like 1
LOC100132600	9	96,234,734	96,235,317	583	hypothetical LOC100132600
FBP2	9	96,360,823	96,395,896	35,073	fructose-1,6-bisphosphatase 2
FBP1	9	96,405,236	96,441,624	36,388	fructose-1,6-bisphosphatase 1
C9orf3	9	96,528,815	96,889,262	360,447	chromosome 9 open reading frame 3
LOC728061	9	96,863,820	96,866,034	2,214	hCG2003663
MIRN23B	9	96,887,310	96,887,406	96	microRNA 23b
MIRN27B	9	96,887,547	96,887,643	96	microRNA 27b
MIRN24-1	9	96,888,123	96,888,190	67	microRNA 24-1
FANCC	9	96,901,157	97,119,812	218,655	Fanconi anemia, complementation group C
LOC643342	9	97,094,441	97,097,142	2,701	similar to ATM/ATR-Substrate Chk2-Interacting Zn2+-finger protein
LOC100132148	9	97,190,452	97,213,751	23,299	hypothetical protein LOC100132148
MT1P1	9	97,215,122	97,215,720	598	metallothionein 1 pseudogene 1
PTCH1	9	97,245,085	97,319,068	73,983	patched homolog 1 (Drosophila)
LOC100130840	9	97,308,338	97,309,739	1,401	hypothetical protein LOC100130840
PSMA7P	9	97,454,864	97,455,469	605	proteasome (prosome, macropain) subunit, alpha type, 7 pseudogene

NCBI36/hg18

SUPPLEMENTARY INFORMATION: Materials and Methods

MSSE patients and DNA extraction

13 MSSE families with 27 affected members included in our study were part of 22 MSSE families members previously described in Goudie *et al*. 2011. DNAs were extracted from either peripheral blood leucocytes or immortalized lymphoblastoid cell lines from patients using standard protocol. Informed consent was obtained from all members of families 2, 4-10, 12, and 15-18 in compliance with local ethical review board requirements. Samples from members of families 1, 3, 11, 13 and 14 were collected with consent for diagnostic testing for Ferguson-Smith disease including haplotype analysis and as such were exempted from specific ethical board review. Normal control DNAs were obtained from Dr. Malcolm Dunlop for Scottish population and CEPH registry. This study was performed with approval of the UCSF Committee On Human Research and conducted according to Declaration of Helsinki Principles.

Targeted next-generation sequencing

RNA oligonucleotides to capture the ~2.2-Mb target region (chr9: 95,253,994-97,455,469 NCBI36/hg18) on chromosome 9q22-31 were designed by e-Array (Agilent Technologies, Santa Clara, CA, USA). Sample libraries were generated according to Agilent SureSelect Target Enrichment kit manual. Amplified captured libraries were sequenced on an Illumina GA II platform (Illumina Inc., San Diego, CA, USA) in UCSF Genome Core Facility. The raw sequencing data were retrieved from the Illumina GA II pipelines.

Sanger dideoxy sequencing

All rare variants listed in Table 1, Tables S2, and S3 were validated by Sanger sequencing (Quintara Biosciences, Albany, CA, USA). The sequencing was performed on both strands. Primer sequences for 9 rare variants shared by Scottish families 2, 4-7, 9, and 18 were designed by Primer 3 (version 0.4.0.) and described in Table 4S.

Electromobility Shift Assay (EMSA) and supershift Assay

EMSA was performed using the Lightshift Chemiluminescent EMSA kit according to manufacturer's instruction (Pierce, Rockford, IL, USA). Biotin-labeled and unlabeled oligonucleotides for *PTCH1* variant 97309311G>C (5'-CTGAATTAGGAAGTGGGGCAGCCAGC-3' and 5'-GCTGGCTGC CCCACTTCCTAATTCAG-3' for wild-type 97309311G; 5'-CTGAATTAGGAAGTGCGGCAGCCAGC-3' and 5'-GCTGGCTGCCGCACTT CCTAATTCAG-3' for variant 97309311C) were synthesized (Sigma, St. Louis, MO, USA). EMSA binding reaction consists of 1X binding buffer, 2.5% glycerol, 5mM MgCl2, 50 ng/uL of poly (dIdC), 0.05% NP40, 7ug of nuclear extract, and 60 fmol of biotin-labeled target DNA. For competition assay, unlabeled target probe was added at 100X and 50X concentration of labeled target probe. The reactions were incubated at room temperature for 20 minutes. For supershift assay, nuclear extract was pre-incubated with 2 μg of SP1 and PU.1 antibodies on ice for 30 minutes. Reactions were electrophoresed on pre-run 6% polyacrylamide gel and transferred onto nylon membrane followed by UV cross-linking. Biotin-labeled target DNA were detected by detection module and exposed to X-ray film.

Bioinformatic analysis

Reads were aligned to the NCBI36/hg18 build of the Human genome using the BWA aligner version 0.5.9 (Li, 2009). Sequences were re-aligned to improve indel calling using the Genome Analysis Toolkit (GATK) (DePristo *et al*., 2011), and variants were called simultaneously using the GATK Universal Genotyper method. Minor allele frequency (MAF) in normal controls was calculated by Haploview 4.2 (Barrett *et al*., 2005).

Barrett JC, Fry B, Maller J *et al*. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-5

DePristo MA, Banks E, Poplin R *et al*. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-8

Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-60