Appendix A

Appendix

VQSR inputs	Variant type	
	SNVs	Indels
Annotations	QD, MQ, MQRankSum, ReadPosRankSum, FS, InbreedingCoeff	QD, FS, ReadPosRankSum, MQRankSum, InbreedingCoeff
Training set	HapMap 3.3: hapmap_3.3.b37.sites.vcf.gz	Mills-Devine and 1000GP Phase i:
	Omni 2.5M chip: 1000G_omni2.5.b37.sites.vcf.gz	Mills_and_1000G_gold_standard.indels.b37.sites.vcf.gz
	1000GP Phase I: 1000G_phase1.snps.high_confidence.b37.vcf	.
Truth set	HapMap 3.3: hapmap_3.3.b37.sites.vcf.gz	Mills-Devine and 1000GP Phase I: Mills_and_1000G_gold_standard.indels.b37.sites.vcf.gz
	Omni 2.5M chip: 1000G_omni2.5.b37.sites.vcf.gz	
Known set	dbSNP build 137: dbsnp_137.b37.excluding_sites_after_129.vcf	dbSNP build 137: dbsnp_138.vcf.gz

Table A.1 Annotations and training sets used for VQSR variant QC.

Variant caller	Filter	Description	Variant type	
			SNPs	Indels
SAMtools	StrandBias	One DNA strand being favored over the other	0.0001	0.0001
	EndDistBias	One DNA strand being favored over the other at the end of reads	0.0001	0.0001
	MaxDP	Maximum depth allowed	2000	2000
	MinDP	Minimum depth allowed	4	4
	MinMQ	Minimum mapping quality allowed	10	10
	MinAB	Minimum number of alternate bases	2	2
	Qual	Minimum value of the overall quality field	10	10
	GapWin	Window size to filter adjacent gaps	3	3
	MapQualBias	Minimum P-value for Mapping quality bias	0	0
	SnpGap	SNP within a certain distance of an Indel to be filtered out	10	10
GATK	MinDP	Minimum depth allowed	4	4
	MinQD	Minimum quality over depth allowed	2	2
	MinMQ	Minimum mapping quality allowed	10	.
	MaxFS	Maximum Fishers P-value	60	200
	MaxHS	Maximum haplotype score	13	.
	MinMQRank	Minimum Z-score	-12.5	.
	MinPosRank	Minimum Z-score	-8	-20
	MinInbreed	Minimum inbreeding coefficient	.	-0.8

Table A.2 Filters and thresholds applied on variants from the TS experiment.

ID	Country of origin	Gender	TSH	fT4 (pmol/L)	Thyroid gland		Comment	Mutations
Monogenic families								
F2a	Saudi Arabia (C)	M	>100	16	Goitre, normal on L-T4 (u)	Siblings	TG R159* (hom)	
F2b	Saudi Arabia (C)	M	150	6.2	Normal on L-T4 (u)		TG R159* (hom)	
F1a	Pakistan (C)	M	>375	3.9	Avid uptake (i)	Siblings	TG R451* (hom)	
F1b	Pakistan (C)	M	NA	NA	NA		TG R451* (hom)	
F3a	Turkish (C)	F	>75	6.4	Nodular, enlarged (u)	Siblings	TG S528* (hom)	
F3b	Turkish (C)	F	>51.4	6	Nodular goitre (u)		TG S528* (hom)	
F4	Oman	M	>100	NA	2 siblings with goitrous CH and the same *TG* genotype	2 affected siblings	TG S2121Afs*32 (hom)	
F5a	UK	M	>150	6	Normal (i)	Siblings	TG R296* (het), TG C160S (het)	
F5b	UK	F	>150	11	NA		TG R296* (het), TG C160S (het)	
F7a	Iraq (C)	M	16 (bs)	NA	Normal (u)	Siblings	TG c.638+5G>A (hom)	
F7b	Iraq (C)	F	43.6	16.8	Avid uptake (i)		TG c.638+5G>A (hom)	
F17	UAE (C)	M	27	NA	Avid uptake, normally sited (i)		TPO R665Q (hom)	
F18	UK	M	920	1.2	Normal (i)		TPO R291H (het), TPO G331V (het)	
F16	UK	F	NA	NA	Nodular goitre (u)		TPO R491H (het), TPO A397Pfs*76 (het)	
F20	Oman (C)	M	7.1 on L-T4	13.5 on L-T4	Avid uptake (i)	2 affected siblings	TPO C808Afs*24 (hom)	
F24	Oman	M	55	NA	Normal (u)		DUOX2 L1028Afs*3 (hom)	
F23	Bangladesh (C)	F	47.6	15.8	Good uptake, normally sited (i)	Transient CH in sibling	DUOX2 F966Sfs*29 (het)	
F26	UAE (C)	M	21	NA	Normal (u)	Cousin with CH	TSHR P68S (het)	
Digenic families								
F8a	Turkey (C)	M	NA	NA	Goitre (u)	Siblings	DUOX2 Q686* (het), TG 1493Y (hom)	
F8b	Turkey (C)	M	79.6	1	Goitre (u)		DUOX2 Q686* (het), TG 1493Y (hom)	
F6a	Turkey (C)	F	123.3	8.9	Nodular goitre (u)	Siblings	TG W1050L (hom), TG C726Y (hom)	
F6b	Turkey (C)	F	NA	NA	Nodular goitre (u)		TG W1050L (hom), TG C726Y (hom), DUOX2 Q686* (het)	
F9a	Turkey (C)	M	>75	TT4 <0.5	Normal (u)	Siblings	TG W2685L (hom), DUOX2 R354W (het)	
F9b	Turkey (C)	M	>75	<1.5	Normal (i)		TG W2685L (hom), DUOX2 R354W (het)	
F11	UAE (C)	M	250	2	Normal (u, b)		TPO R491H (hom), TG Q1644E (het)	
F21	UK	M	>100	3.8	Avid uptake, enlarged (l, u)		TPO E17Dfs*77 (het), TPO Y453D (het), SLC26A4 E384G (het)	
F10a	Pakistan (C)	M	>150	NA	Normal (u)	Cousins with transient CH	DUOX2 Q570L (hom), TG R1691C (het), TG L2647Q (het)	
F19a	UK/African-Caribbean	F	400	0.8	Normal (i)	Siblings	TPO R584Q (hom), SLC26A4 N324Y (het)	
F19b	UK/African-Caribbean	M	620	0.7	Normal (i)		TPO R584Q (hom), SLC26A4 I713M (het)	

Table A.3 Detailed genotype and phenotype information for all CH patients with causative mutations in the eight known *gland-in-situ* genes.

ID	Country of origin	Gender	TSH	fT4 (pmol/L)	Thyroid gland	Other features	Mutations
Ambiguous cases							
F22	UK	F	10	14	Normal (u)		TPO E510Afs*14 (het)
F12a	India/UK	M	11.3	13.6	Left lobe smaller than right (u)	Siblings	TG Q870H (het)
F12b	India/UK	M	10.05	12.3	Left lobe smaller than right (u)		TG Q870H (het)
F13	Wales	M	96	12	Normal (u)	Cleft palate	TG Q771* (het)
F15a	Somalia	F	Mildly elevated	NA	Normal (u)	Siblings	TG c.3433+3_3433+6delGAGT (het)
F15b	Somalia	M	Mildly elevated	NA	Normal (u)		TG c.3433+3_3433+6delGAGT (het)
F15c	Somalia	M	40.3	17.8	Avid uptake (i)		TG c.3433+3_3433+6delGAGT (het)
F14a	Yemen (C)	F	NA	NA	Normal (u)	Identical twins	TG Y759C (het)
F14b	Yemen (C)	F	8	16	Normal (u)		TG Y759C (het)
F25*	Sri Lanka	M	NA	NA	Avid uptake (i)	2 affected siblings	DUOX2 R764W (het)
F27	Bangladesh (C)	F	>50	4.3	Normal (u)		DUOXA2 c. 555-5G>A (hom)
Unsolved cases							
F28	UAE (C)	M	200	NA	Normal (u)		
F29a	Saudi Arabia (C)	M	>100	1.4	Goitre	Siblings	
F29b	Saudi Arabia (C)	M	>100	0.49	NA		
F30	Oman (C)	M	NA	NA	Normal (u)	1 affected sibling	
F31	Australia	M	20.7	11.5	Normally sited, decreased uptake (i)	1 affected sibling	
F32	Poland	F	31.2	12.4	Normal (i)		
F33a	UK	F	12.3 (bs)	NA	Normal (i)	Siblings	
F33b	UK	F	22	NA	NA		
F34	UAE (C)	M	153	5.7	Normally sized and sited (u)		

Table A.4 Detailed genotype and phenotype information for all CH patients that were considered 'ambiguous' or 'unsolved' due to a lack of convincing causative variants in *gland-in-situ* genes.

Mouse gene	Human gene
Mouse models of TD	
Shh	*SHH*
Foxe1	*FOXE1*
Chrd	*CHRD*
Edn1	*EDN1*
Eya1	*EYA1*
Fbln1	*FBLN1*
Hes1	*HES1*
Hoxa5	*HOXA5*
Isl1	*ISL1*
Nkx2-5	*NKX2-5*
Frs2	*FRS2*
Hoxa3	*HOXA3*
Hoxb3	*HOXB3*
Hoxd3	*HOXD3*
Pax3	*PAX3*
Fgfr2	*FGFR2*
Fgf10	*FGF10*
Hhex	*HHEX*
Nkx2-1	*NKX2-1*
Pax8	*PAX8*
Twsg1	*TWSG1*
Tbx1	*TBX1*
Zebrafish models of TD	
ace	*ACE*
cyc	*CYC1*
fau	*FAU*
hand2	*HAND2*

Mouse gene	Human gene
Genes enriched in the mouse thyroid bud at E10.5	
Adrbk2	*ADRBK2*
Afap1L2	*AFAP1L2*
Atp10a	*ATP10A*
Bcl11b	*BCL11B*
Bcl2	*BCL2*
Calml3	*CALML3*
Capg	*CAPG*
Cckar	*CCKAR*
Cd44	*CD44*
Chdh	*CHDH*
Clstn2	*CLSTN2*
Cpne4	*CPNE4*
Cpxm2	*CPXM2*
Ctxn3	*CTXN3*
Cxcl12	*CXCL12*
Elfn1	*ELFN1*
Galns	*GALNS*
Gcgr	*GCGR*
Hexb	*HEXB*
Hivep3	*HIVEP3*
Htra1	*HTRA1*
Irs4	*IRS4*
Klhl14	*KLHL14*
Lypd6b	*LYPD6B*
Matn2	*MATN2*
Nbeal2	*NBEAL2*
Nptx1	*NPTX1*
Pla2g7	*PLA2G7*
Prlr	*PRLR*
Ptpre	*PTPRE*
Ryr3	*RYR3*
Scara5	*SCARA5*
Slc16a2	*SLC16A2*
Slc44a3	*SLC44A3*
Slc4a4	*SLC4A4*
Slc4A5	*SLC4A5*
Sorbs2	*SORBS2*
Stc2	*STC2*
Tbx3	*TBX3*
Tcfcp2l1	*TFCP2L1*
Zbtb20	*ZBTB20*
Zbtb4	*ZBTB4*

Table A.5 List of CH candidate genes, part 1.

Mouse gene	Human gene
Genes enriched in both the mouse thyroid bud and lung at E10.5	
Alcam	*ALCAM*
Ap1m2	*AP1M2*
Arhgef16	*ARHGEF16*
Cdcp1	*CDCP1*
Cdh1	*CDH1*
Cdh16	*CDH16*
Cldn3	*CLDN3*
Cldn6	*CLDN6*
Cldn7	*CLDN7*
Clu	*CLU*
Crb3	*CRB3*
Ct14	*SAGE1*
Dsg2	*DSG2*
Epcam	*EPCAM*
Eppk1	*EPPK1*
Esrp2	*ESRP2*
Inadl	*INADL*
Kcnk1	*KCNK1*
Mapk13	*MAPK13*
Marveld2	*MARVELD2*
Marveld3	*MARVELD3*
Mbip	*MBIP*
Meg3	*MEG3*
Mfsd6	*MFSD6*
Npnt	*NPNT*
Pitpnm1	*PITPNM1*
Prss8	*PRSS8*
Pygl	*PYGL*
Rassf10	*RASSF10*
Ripk4	*RIPK4*
Sh3gl2	*SH3GL2*
Slco2a1	*SLCO2A1*
Sorcs2	*SORCS2*
Sorl1	*SORL1*
Spint1	*SPINT1*
Spint2	*SPINT2*
Tie2	*TEK*
Tmem176a	*TMEM176A*
Tnk1	*TNK1*

Foxe1 targets	Pax8 targets
AHCY	*CDH16*
AMIGO3	*CITED2*
ANKRD37	*EGR1*
ATMIN	*IGFBP7*
BET1	*KCNJ15*
CASP4	*KCNJ16*
COQ10B	*NFKB1*
CRELD2	*RAB17*
CTGF	*RUNX2*
DDIT3	*SPARC*
DERL3	*TRIB1*
DNAJB11	*WBP2*
DNAJB9	*WNT4*
DNAJC3	
DUSP5	
ENGASE	
ERO1LB	
ETV5	
GGCT	
GMPPB	
HSP90B1	
HSPA5	
HYOU1	
IGF2BP2	
IL23A	
MANF	
MFSD2	
NR4A2	
NUPR1	
PDIA4	
RIOK3	
SDF2L1	
SEC23B	
SEL1L	
TM4SF1	
TMEM66	
ZFAND2A	
ADAMTS9	
BCAM	
CDH1	
CRIP2	
DYNLRB2	
ELOVL2	
FGF18	
FOLR1	
KRT20	
PRIMA1	
PRSS8	
RIL	
S100A4	
SLIT1	
TMEM140	

Table A.6 List of CH candidate genes, part 2.

Baseline patient characteristics		N = 145	
Gender (*female/male*)		63 (43%) / 82 (57%)	
Ethnicity (*Caucasian / Black / Asian / Jewish / Others / unknown)*		99 (68%) / 2 / 21 / 1 / 11 / 11	
Age at diagnosis, years			
Mean ± SD		3.6 ± 1.8	
Median (range)		3.5 (range 4 weeks to 7 years)	
Diagnosis (*CD / UC / IBDU*)		66 (46%) / 51 (35%) / 28 (19%)	
Positive family history		29/137 (21.2%)	
Paris Crohn's Classification (n=66) *			
Disease location	L1 Ileum	2	
	L2 Colon	20	
	L3 Ileocolonic	42	
	+ L4 (upper GI tract)	34	
Disease behavior	B1 nonstricturing-nonpenetrating	54	
	B2 sticturing	6	
	B3 penetrating	2	
	B2B3 penetrating and stricturing	3	
	Perianal involvement	17 (24.2%)	
Paris UC classification		**UC (n=49) ****	**IBDU (n=25) *****
Disease location	E1 Ulcerative proctitis	1	0
	E2 Left sided UC to splenic flexure	6	0
	E3 Extensive UC to hepatic flexure	4	1
	E4 Pancolitis	38	24
GI surgery			
	Colectomy	23/144 (16.0%)	
Associate medical therapy			
	Steroids	93/144 (64.6%)	
	Azathioprine/6-MP	102/144 (70.8%)	
	MTX	11/144 (7.6%)	
	CYA	9/144 (6.3%)	
	Anti-TNFα	51/144 (35.4%)	

Table kindly generated and provided by Dr Tobias Schwerd.
One data set incomplete (David Wilson), "unknown" patient details excluded for analysis;

* CD: 2/66 patients with oral and perianal CD only
** UC: 1/51 patient unknown location, 1/51 data set incomplete
*** IBDU: 3/28 patients without macroscopic inflammation

Table A.7 Disease phenotype and therapy characteristics for the VEO-IBD cohort.