

1 ***ESR1, WT1, WNT4, ATM and TERT* loci are major contributors to uterine leiomyoma**
2 **predisposition**

3 Niko Välimäki†¹, Ph.D., Heli Kuisma†¹, M.Sc., Annukka Pasanen², M.D., Oskari
4 Heikinheimo³, M.D., Ph.D., Jari Sjöberg³, M.D., Ph.D., Ralf Bützow², M.D., Ph.D., Nanna
5 Sarvilinna^{1,3,4}, M.D., Ph.D., Hanna-Riikka Heinonen¹, M.D., Ph.D., Jaana Tolvanen¹, M.Sc.,
6 Simona Bramante¹, Ph.D., Tomas Tanskanen¹, M.D., Ph.D., Juha Auvinen⁵, M.D., Ph.D.,
7 Terhi Piltonen⁶, M.D., Ph.D., Amjad Alkodsı¹, Ph.D., Rainer Lehtonen¹, Ph.D., Eevi
8 Kaasinen⁷, Ph.D., Kimmo Palin¹, Ph.D., and Lauri A. Aaltonen^{1*}, M.D., Ph.D.

9 ¹Department of Medical and Clinical Genetics and Genome-Scale Biology Research Program,
10 University of Helsinki, Helsinki, Finland; ²Department of Pathology, University of Helsinki and
11 Helsinki University Hospital, Helsinki, Finland; ³Department of Obstetrics and Gynecology,
12 University of Helsinki and Helsinki University Hospital, Helsinki, Finland; ⁴Institute of Biomedicine,
13 Biochemistry and Developmental Biology, University of Helsinki, Helsinki, Finland; ⁵Northern
14 Finland Birth Cohort Studies and Center for Life Course Health Research, Faculty of Medicine,
15 University of Oulu, Oulu, Finland; ⁶Department of Obstetrics and Gynecology, PEDEGO Research
16 Unit, Medical Research Center Oulu, Oulu University Hospital, University of Oulu, Oulu, Finland;
17 ⁷Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and
18 Biophysics, Karolinska Institutet, Stockholm, Sweden

19 † These authors contributed equally to this work.

20 * Correspondence to: Lauri A. Aaltonen, Department of Medical and Clinical Genetics,
21 Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), FI-00014 University of Helsinki,
22 Finland. E-mail: lauri.aaltonen@helsinki.fi

1 **ABSTRACT**

2 Uterine leiomyomas (ULs) are benign tumors that are a major burden to women's health. A
3 genome-wide association study on 5,417 UL cases and 331,791 controls was performed,
4 followed by replication of the genomic risk in two cohorts. Effects of the identified risk
5 alleles were evaluated in view of molecular and clinical features.
6 Five loci displayed a genome-wide significant association; the previously reported *TNRC6B*,
7 and four novel loci *ESR1* (*ER α*), *WT1*, *WNT4*, and *ATM*. The sixth hit *TERT* is also a
8 conceivable target. The combined polygenic risk contributed by these loci was associated
9 with *MED12* mutation-positive tumors. The findings link genes for uterine development and
10 genetic stability to leiomyomagenesis. While the fundamental role of sex hormones in UL
11 aetiology has been clear, this work reveals a connection to estrogen receptor alpha on genetic
12 level and suggests that determinants of UL growth associated with estrogen exposure have an
13 inherited component.

1 INTRODUCTION

2 Uterine leiomyomas (ULs), also known as fibroids or myomas, are benign smooth muscle
3 tumors of the uterine wall. They are extremely common; approximately 70% of women
4 develop ULs before menopause¹. The symptoms, occurring in one fifth of women, include
5 excessive menstrual bleeding, abdominal pain and pregnancy complications¹. In most cases,
6 durable treatment options are invasive². ULs cause a substantial human and economic
7 burden, and the annual cost of treating these tumors has been approximated to be as high as
8 \$34 billion in the United States, higher than the combined cost of treating breast and colon
9 cancer³.

10 Earlier studies have indicated strong genetic influence in UL susceptibility based on linkage⁴,
11 population disparity⁵ and twin studies⁶. The most striking UL predisposing condition thus far
12 characterized is hereditary leiomyomatosis and renal cell cancer (HLRCC) syndrome, caused
13 by high-penetrance germline mutations in the *Fumarate hydratase (FH)* gene^{7,8}.
14 Genome-wide association studies (GWAS) have proposed several low-penetrance risk loci
15 but few unambiguous culprit genes have emerged. Cha *et al.* reported loci in chromosome
16 regions 10q24.33, 11p15.5 and 22q13.1 based on a Japanese patient cohort⁹. The 11p15.5
17 locus - near the *Bet1 golgi vesicular membrane trafficking protein like (BET1L)* gene - was
18 later replicated in European Americans¹⁰. The 22q13.1 locus has been replicated in
19 Caucasian, American and Saudi Arabian populations suggesting *trinucleotide repeat*
20 *containing 6B (TNRC6B)* as a possible target gene¹⁰⁻¹². Further UL predisposition loci have
21 been suggested at 1q42.2 and 2q32.2 by Zhang *et al.*¹³ and, at 3p21.31, 10p11.21 and 17q25.3
22 by Eggert *et al.*¹⁴ A recent work reported *cytohesin 4 (CYTH4)* at 22q13.1 as a novel

1 candidate locus in African Americans¹⁵. While multiple loci and genes have been implicated
2 through these valuable studies it is not straightforward to connect any of them
3 mechanistically to UL development.

4 Most ULs show somatic site-specific mutations at exons 1 and 2 of the *mediator complex*
5 *subunit 12 (MED12)* gene^{16,17}. These observations together with further scrutiny of driver
6 mutations, chromosomal aberrations, gene expression, and clinicopathological characteristics
7 have lead to identification of at least three mutually exclusive UL subtypes; *MED12* mutant,
8 *Fumarate Hydratase* deficient, as well as *HMGA2* overexpressing lesions¹⁸.

9 Here we report the most powerful GWAS on uterine leiomyoma to date, and novel
10 genome-wide significant UL susceptibility loci with plausible adjacent culprit genes
11 including *Estrogen Receptor 1 (ESR1 or ER α)*, *Wilms Tumor 1 (WT1)*, *Wnt Family Member 4*
12 (*WNT4*) and *ATM Serine/Threonine Kinase (ATM)*. Genome-wide significant associations
13 were observed also at the previously reported locus in *TNRC6B*, and after meta-analysis at a
14 gene poor locus 13q14.11. *Telomerase Reverse Transcriptase (TERT)* locus narrowly missed
15 genome-wide significance, but *TERT* appeared as a highly plausible UL predisposition gene
16 supported by overwhelming evidence in many other tumor types¹⁹. We compiled the weight
17 of these susceptibility loci into a polygenic risk score and replicated the UL association in an
18 independent cohort of Finnish origin. Finally, we investigated the risk alleles' association to
19 clinical features, molecular UL subtypes, telomere length, gene expression and DNA
20 methylation.

1 **METHODS**

2 **Genome-wide association study**

3 Fig. 1 provides an outline of the four stages that were implemented. The discovery stage
4 GWAS was based on UK Biobank (UKBB) genotypes: a Global Biobank Engine (GBE)
5 query was used to test association between 5,417 UL cases and 331,791 controls of
6 Caucasian ancestry. The second stage meta-analysis utilized the genome-wide summary
7 statistics from UKBB and the Helsinki cohort of 457 UL cases and 15,943 controls. The
8 observations were replicated using 459 UL cases and 4,943 controls from the Northern
9 Finland Birth Cohort (NFBC). More details on the GWAS materials and methods are given in
10 Supplementary Methods.

11 **Patient and tumor material**

12 Our in-house patient and tumor data were investigated regarding the identified risk loci.
13 Clinical background data for the number of ULs (Fig. S1) and age at hysterectomy (Fig. S2)
14 were available for 357 patients in the Helsinki cohort. All tumors of ≥ 1 cm diameter had been
15 harvested and stored fresh-frozen, and *MED12*-mutation status was screened from all 1,165
16 tumors. Gene expression was studied in a 1Mb flank from each SNP of interest using 60
17 tumors and 56 respective normal myometrium samples. DNA methylation was studied in an
18 1Mb flank using 56 tumors and 36 normals. Telomere length estimates derived from whole
19 genome sequencing data were available for 37 UL and 28 myometrium samples. The study

1 was approved by the ethics review board of the Helsinki University Central Hospital,
2 Finland. Details regarding materials and methods are given in the Supplementary Methods.

3 **Statistical analysis**

4 For GWAS, $P < 5 \times 10^{-8}$ was reported as significant. The GRS association tests (Table S6) were
5 controlled for family-wise error rate (FWER) and reported significant for Holm-Bonferroni
6 adjusted $P < 0.05$. Other families of association tests were controlled for false discovery rate
7 (FDR; Benjamini-Hochberg method) and noted significant at $FDR < 10\%$.

8 **RESULTS**

9 **Identification of predisposition loci**

10 At discovery stage 11 SNPs emerging from five distinct genetic loci passed the genome-wide
11 significance level of 5×10^{-8} : 11q22.3 region in *ATM* (top SNP rs1800057; OR=1.49;
12 $P < 1 \times 10^{-16}$), 11p13 near *WT1* (rs74911261; OR=1.21; $P = 1.7 \times 10^{-13}$), 6q25.2 near *ESR1*
13 (rs11751190; OR=1.17; $P = 1.4 \times 10^{-11}$), 1p36.12 near *WNT4* (rs12038474; OR=1.17;
14 $P = 1.8 \times 10^{-10}$) and at 22q13.1 in *TNRC6B* (rs12484776; OR=1.14; $P = 2.5 \times 10^{-8}$). The sixth most
15 significant association, not reaching genome-wide significance, was observed at *TERT* locus
16 in 5p15.33 (rs2736100; OR=0.91; $P = 3.6 \times 10^{-7}$). The association results are given in Table S1,
17 and the lead SNPs are summarized in Table 1.

1 Fig. 2 displays a regional structure of each locus and the flanking association values, linkage
2 disequilibrium (LD) and genome annotation. Annotation tracks are included for
3 tissue-specific data on open chromatin, topologically associating domains (TAD) and other
4 regulatory features (details in Supplementary Methods).

5 **Genomic risk score**

6 A polygenic risk score²⁰ was compiled based on the discovery stage associations, including
7 *TERT* locus. After LD thinning ($r^2 < 0.3$) the 12 discovery-stage SNPs, eight SNPs from the six
8 distinct loci passed for the initial genomic risk score (GRS; Table S3). The SNP weights were
9 based on UKBB log-odds. We applied this initial GRS of eight SNPs to the Helsinki cohort
10 and identified a significant association to the UL phenotype (Wilcoxon rank-sum $P=0.0098$;
11 adjusted $P=0.029$; one-sided; $W=3.41 \times 10^6$).

12 **Meta-analysis**

13 The second stage GWAS combined the UKBB and Helsinki cohorts for a meta-analysis
14 approach. We utilized the genome-wide statistics available for a total of 5,168 self-reported
15 UL cases (Neale lab data; details in Supplementary Methods). rs117245733 at 13q14.11 was
16 identified as the only SNP with a suggestive ($P < 10^{-6}$) association in both the UKBB
17 (OR=1.34; $P=3.31 \times 10^{-6}$) and Helsinki (OR=1.82; $P=8.12 \times 10^{-6}$) cohorts. The combined
18 association was genome-wide significant (fixed effect model OR=1.41; $P=3.20 \times 10^{-8}$). Fig. 3
19 shows the regional structure and combined association at the locus: the SNP resides on a gene
20 poor region, at a conserved element that has activity in uterus-specific H3K27ac and DNaseI

1 data (see ENCODE track details in Supplementary Methods). Summary of the meta-analysis
2 results is given in Table S2. The SNP rs117245733 at 13q14.11 - weighted by its UKBB
3 log-odds - was appended to the initial GRS model. The final GRS model of nine SNPs and
4 their UKBB-based weights is given in Table S3.

5 **Replication of the GWAS and GRS**

6 The third stage replicated the observations in NFBC. The SNP identified in the stage 2
7 meta-analysis, rs117245733 at 13q14.11, was replicated with a nominal $P=0.034$ (linear
8 mixed model; $OR=1.50$; 95% CI 1.03-2.19). All SNPs except rs11031736 at 11p13 had the
9 same effect direction as observed in UKBB, however, none of the single-SNP associations in
10 NFBC held with multiple testing (Table S4). The association between the GRS and UL
11 phenotype replicated with a significant $P=1.7\times 10^{-5}$ (Wilcoxon rank-sum; adjusted $P=0.0001$;
12 one-sided; $W=1.0\times 10^6$). Analysis of the case-control distributions showed an odds ratio of
13 2.09 for one-unit increase in GRS values (logistic regression; 95% CI 1.43-3.04; Fig. S4-S5).
14 The GRS model passed the goodness-of-fit test (Hosmer-Lemeshow $P=0.406$), and the
15 sensitivity and specificity characteristics of the model are summarized as a receiver operating
16 characteristic (ROC; $AUC=0.56$; 95% CI 0.53-0.59) curve in Fig. S5B.

17 UL susceptibility is known to vary by ancestry⁵. We computed population specific estimates
18 for the GRS based on the gnomAD database allele frequencies. Fig. S5 shows the resulting
19 estimates for reference (details in Table S7).

1 **Association to clinical variables**

2 The number of ULs per patient had a significant positive association to the GRS (negative
3 binomial regression $P=0.0047$; adjusted $P=0.024$; rate ratio 1.61; 95% CI 1.07-2.44 for
4 one-unit increase in GRS; Fig. S6). No association was found between the GRS and age at
5 hysterectomy (Table S6). Tests for each SNP separately are given in Table S5.

6 **Association to *MED12* mutated tumors**

7 Our UL set of 1165 lesions included 931 (80%) mutation-positive and 234 mutation-negative
8 tumors. the occurrence of mutant tumors did not distribute evenly among the 357 patients. In
9 total 178 (50%) and 96 (27%) patients had all their tumors identified as either
10 *MED12*-mutation-positive or -negative, respectively (Fig. S7), suggesting that genetic or
11 environmental factors contribute to the preferred UL type in an affected individuals, as
12 previously observed¹⁶. Indeed, mutation positive patients were found to have a significantly
13 higher GRS (Wilcoxon rank-sum $P=0.005$; adjusted $P=0.024$; two-sided; $W=6803$). This
14 difference in GRS distributions is visualized in Fig. S5.

15 Strikingly, comparison against the population controls ($n=15,943$) revealed an opposite effect
16 direction for the above-mentioned patient groups: the *MED12*-mutation-positive (178) subset
17 of patients displayed an odds ratio of 2.34 for one-unit increase in GRS (95% CI 1.32-4.15)
18 compared to the controls, and the mutation-negative (96) subset an odds ratio of 0.66 (95%
19 CI 0.28-1.54) compared to the controls. Thus, the majority of the compiled case-control
20 association signal had arisen from the *MED12*-mutation-positive subset of the patients.

1 The number of *MED12*-mutation-positive tumors per patient also had a significant positive
2 association to the GRS (negative binomial model $P=0.0026$; adjusted $P=0.018$; rate ratio
3 2.08; 95% CI 1.13-3.83 for one-unit increase in GRS; Fig. S8). No association between the
4 number of *MED12*-mutation-negative tumors and GRS was found (Fig. S8). Tests for each
5 SNP separately are given in Table S5.

6 **Association to gene expression**

7 In tumors, one cis expression quantitative trait loci (cis-eQTLs) passed a 10% FDR. The risk
8 allele of rs61778046 at 1p36.12 showed positive correlation with both *WNT4* expression
9 (nominal $P=0.004$, Fig. 4B) and *alkaline phosphatase, liver/bone/kidney (ALPL)* expression
10 ($P=3.07\times 10^{-5}$, Fig. S11). *WNT4* and *ALPL* are in adjacent TADs (Fig. 4C). In normal
11 myometria, no cis-eQTL passed the 10% FDR. Altogether 11 genes in tumors and 8 genes in
12 normals passed with a nominal $P<0.05$ (Table S8). All expression results within 1Mb flank
13 can be found in Supplementary Data 1-2.

14 Altogether 95 cis-splicing quantitative trait loci (cis-sQTL) had a suggestive correlation ($P <$
15 0.05) within 1Mb flank of each SNP. No cis-sQTL passed the 10% FDR. All the cis-sQTLs
16 are listed in Supplementary Data 4.

17 **Association to DNA methylation**

18 Altogether 3,340 (1,865 in tumors and 1,475 in normals) cis methylation quantitative trait
19 loci (cis-meQTL) had suggestive associations with nominal $P < 0.05$. Of these, 38 passed a

1 10% FDR. Of the plausible culprit genes, *TERT*, *WT1* and *WNT4* showed most significant
2 meQTL associations. 18 of the meQTLs in these genes were detectable both in tumors and
3 normals (Table S9). All the cis-meQTLs and annotation for their genomic context are in
4 Supplementary Data 3. The association of selected meQTLs with expression is shown in Fig.
5 S12-S13.

6 **Association to telomere length**

7 When examining telomere lengths in tumors from carriers and noncarriers of the *TERT* risk
8 allele, the risk allele (rs2736100) was significantly associated with a shorter telomere length
9 (Kruskal-Wallis $P=0.01$) (Fig. S14). The association was not seen in myometrium. Overall
10 the telomere length was significantly shorter in tumors than normals ($P=0.01$), as previously
11 reported^{21,22}. Adjusting for the patient age did not explain away the association in tumor data.
12 No association was detected between the genotype and the number of somatic structural
13 variants.

14 **Previously proposed UL predisposition loci**

15 Previous UL association studies^{9,13-15} have reported altogether seven genome-wide significant
16 UL susceptibility loci. Two out of the seven loci - that is, 22q13.1 (at *TNRC6B*) and 11p15.5
17 (at *BET1L*) - replicated in the UKBB summary statistics for 5,168 self-reported UL cases
18 (Neale lab data). See Table S10 for a summary of these results.

1 DISCUSSION

2 The UK Biobank genotype-phenotype data revealed five novel predisposition loci for UL in
3 close proximity of highly plausible culprit genes. A meta-analysis of the UKBB and Helsinki
4 cohorts identified a sixth novel locus at 13q14.11, residing within a conserved element that
5 has enhancer activity in uterus-specific experiments (Fig. 3). Two previously reported loci, at
6 *TNRC6B*⁹⁻¹² and *BET1L*^{9,23}, were also validated, and are indeed linked to UL predisposition.
7 For the three latter loci mechanistic connection to UL is obscure at present.

8 Though simple association is not sufficient to formally prove causality, the five new risk
9 genes implicated in the discovery phase of this study, *ESR1*, *WT1*, *WNT4*, *ATM* and *TERT*,
10 form a robust set of likely culprits.

11 Estrogen is a well known inducer of UL growth²⁴. The top association at 6q25.2 (rs11751190)
12 resides within intron 23 of *Spectrin Repeat Containing Nuclear Envelope Protein 1* (*SYNE1*),
13 130kb downstream of *ESR1*, the latter being the only gene that resides completely within the
14 TAD (Fig. 2C). While the role of estrogen in leiomyomagenesis has been firmly established,
15 this is the first genetic evidence to this end.

16 *WT1* and *WNT4* are central factors in uterine development^{25,26}, and perturbations in their
17 function are known to have neoplastic potential. The strongest association at 11p13
18 (rs2057178) maps to an intergenic region, 45kb downstream of the closest gene *WT1* and the
19 region has active enhancer activity in uterus (Fig. 2B). *WT1* is a transcription factor that acts
20 as both a tumor suppressor and an oncogene²⁷. The chromosome 1p36.12 association

1 (rs12038474 and rs61778046) arises at the promoter region of *Cell Division Cycle 42*
2 (*CDC42*) and roughly 50kb downstream of *WNT4*. While *CDC42* cannot be dismissed as a
3 candidate gene, tumor derived data suggests *WNT4* as the more plausible target. The risk
4 allele at rs61778046 was associated with upregulation of *WNT4* (Fig. 4B). The effect is also
5 supported by GTEx Portal data, where *WNT4* shows suggestive upregulation associated with
6 the presence of risk allele in myometrium (P=0.044; accessed Dec 1, 2017). *WNT4* is known
7 to be overexpressed in uterine leiomyomas with *MED12* mutations²⁸. Taken together the risk
8 allele may cause increased expression already in the normal myometrium, and the effect
9 could be further selected for during UL genesis. The risk locus in 1p36.12 was also associated
10 with several meQTLs suggesting that methylation may have a role in *WNT4* regulation (Fig.
11 4). *WNT4* encodes a signaling protein that has a crucial role in sex-determination²⁹, and the
12 WNT signaling pathway has a well-established role in various malignancies such as breast
13 and ovarian cancer³⁰. Of note, recent GWAS on gestational duration suggested that binding of
14 the estrogen receptor at *WNT4* is altered by rs3820282 ($r^2=0.9$ with our lead SNP)³¹.

15 *ATM* and *TERT* could be involved in uterine neoplasia predisposition through genetic
16 instability. The lead SNP at 11q22.3 is a nonsynonymous variant (rs1800057; p.Pro1054Arg)
17 in exon 22 of *ATM*. This SNP was recently highlighted in a renal cell carcinoma GWAS³².
18 *ATM* is involved in DNA damage response³³ and is one of the relatively few genes that have
19 been found to be recurrently mutated in leiomyosarcoma³⁴. *TERT* encodes a subunit of the
20 telomerase enzyme, which guards chromosomal stability by elongating telomeres. It is
21 expressed in germ cells as well as in many types of cancers³⁵. The risk allele at the *TERT*
22 intron 2 (5p15.33; rs2736100) has several associated cis-meQTLs in the 1Mb flank region
23 (Fig. S9-S10). For example CpG chr5:1277576 in the sixth intron of *TERT* and CpG

1 chr5:1285974 in the second intron of *TERT* are detectable both in tumors and normals with
2 nominal $P < 0.001$. While the neoplasia predisposing effect of this SNP is overwhelmingly
3 documented, previous studies have reported contradicting observations on its effect on
4 telomere length^{36–40}. ULs have been shown to display shortened telomeres^{21,22}, potentially
5 provoking chromosomal instability as the chromosome tips are worn out. Our data shows that
6 this finding is explained by *TERT* risk allele carrier status (Fig. S14).

7 GRS associated merely with a susceptibility to the most common UL subtype, *MED12*
8 mutation positive tumors. Indeed it has been known that *MED12*-mutation-positive tumors do
9 not distribute randomly among patients¹⁶, and our data provide at least a partial explanation to
10 this intriguing finding. It may be that environmental factors contribute more significantly to
11 genesis of *MED12* wild type lesions. In our recent study this tumor type was associated with
12 history of pelvic inflammatory disease, and thus infectious agents could be one underlying
13 factor⁴¹. Obviously, also the power of GWAS to detect genetic associations to more rare UL
14 types is reduced. Much additional work is needed to fully elucidate the mechanisms
15 connecting the risk alleles and emergence of *MED12* mutant UL.

16 This work highlights several new genetic cornerstones of UL formation, and represents
17 another step towards a much improved understanding of its molecular basis. The GRS score
18 can stratify the female population to low-risk and high-risk quartiles that differ two-fold by
19 their UL risk. While the increased risk appears minor on individual level, the population-level
20 burden to women's health arising from these risk loci is highly significant considering the
21 incidence of the condition. Together with the recent progress in molecular tumor
22 characterization and subclassification, the identification of the genetic components of UL

1 predisposition should pave the way towards more sophisticated prevention and management
2 strategies for these extremely common tumors. The risk SNP with the most immediate
3 potential value is that at estrogen receptor alpha, and our findings should fuel much further
4 work on the interplay between individual germline genetics, endogenous and exogenous
5 hormonal exposure, and occurrence and growth rate of UL.

6 **ACKNOWLEDGEMENTS**

7 We are thankful to Sini Marttinen, Sirpa Soisalo, Marjo Rajalaakso, Inga-Lill Åberg, Iina
8 Vuoristo, Alison Ollikainen, Elina Pörsti, Salla Välipakka and Heikki Metsola for their
9 technical support. We also thank Pirjo Ikonen and the rest of the staff of the Kätilöopisto
10 Maternity Hospital, and the staff of the Department of Pathology, University of Helsinki for
11 technical assistance. We thank Minna Männikkö, Tuula Ylitalo, and the rest of the staff of the
12 Northern Finland Birth Cohort Studies, Faculty of Medicine, University of Oulu for technical
13 assistance. The study was supported by grants from Academy of Finland (Finnish Center of
14 Excellence Program 2012-2017, No. 1250345), European Research Council (ERC, 695727),
15 Cancer Society of Finland, Sigrid Juselius Foundation and Jane and Aatos Erkko Foundation.
16 NV received a grant from the Academy of Finland (No. 287665). KP received a grant from
17 Nordic Information for Action eScience Center (NIASC), the Nordic Center of Excellence
18 financed by NordForsk (No. 62721). The authors would like to thank the Rivas lab for
19 making the GBE resource available. We thank the Neale lab for making the UKBB summary
20 statistics available.

1 AUTHOR CONTRIBUTIONS

2 AP, OH, JS, RB and NS collected the tissue samples. NV and HK analyzed the data and
3 wrote the first draft of the manuscript. HRH, JT, SB, TT, JA, TP, AA, RL and EK participated
4 in the data collection and analysis. LAA and KP supervised the study. All authors reviewed
5 the final draft.

6 COMPETING FINANCIAL INTERESTS

7 The authors declare no competing financial interests.

8 REFERENCES

- 9
- 10 1. Stewart, E. A., Cookson, C. L., Gandolfo, R. A. & Schulze-Rath, R. Epidemiology of uterine
11 fibroids: a systematic review. *BJOG* **124**, 1501–1512 (2017).
 - 12 2. Stewart, E. A. Clinical practice. Uterine fibroids. *N. Engl. J. Med.* **372**, 1646–1655 (2015).
 - 13 3. Cardozo, E. R., Clark, A. D., Banks, N. K., Henne, M. B., Stegmann, B. J. & Segars, J. H. The
14 estimated annual cost of uterine leiomyomata in the United States. *Am. J. Obstet. Gynecol.* **206**,
15 211.e1–9 (2012).
 - 16 4. Gross, K. Finding genes for uterine fibroids. *Obstetrics & Gynecology* **95**, S60 (2000).
 - 17 5. Wise, L. A., Ruiz-Narvaez, E. A., Palmer, J. R., Cozier, Y. C., Tandon, A., Patterson, N., Radin,
18 R. G., Rosenberg, L. & Reich, D. African ancestry and genetic risk for uterine leiomyomata. *Am.*
19 *J. Epidemiol.* **176**, 1159–1168 (2012).
 - 20 6. Luoto, R., Kaprio, J., Rutanen, E. M., Taipale, P., Perola, M. & Koskenvuo, M. Heritability and
risk factors of uterine fibroids--the Finnish Twin Cohort study. *Maturitas* **37**, 15–26 (2000).

- 1 7. Tomlinson, I. P. M., Alam, N. A., Rowan, A. J., Barclay, E., Jaeger, E. E. M., Kelsell, D., Leigh,
2 I., Gorman, P., Lamlum, H., Rahman, S., Roylance, R. R., Olpin, S., Bevan, S., Barker, K.,
3 Hearle, N., Houlston, R. S., Kiuru, M., Lehtonen, R., Karhu, A., Vilkki, S., Laiho, P., Eklund, C.,
4 Vierimaa, O., Aittomäki, K., Hietala, M., Sistonen, P., Paetau, A., Salovaara, R., Herva, R.,
5 Launonen, V., Aaltonen, L. A. & Multiple Leiomyoma Consortium. Germline mutations in FH
6 predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell
7 cancer. *Nat. Genet.* **30**, 406–410 (2002).
- 8 8. Launonen, V., Vierimaa, O., Kiuru, M., Isola, J., Roth, S., Pukkala, E., Sistonen, P., Herva, R. &
9 Aaltonen, L. A. Inherited susceptibility to uterine leiomyomas and renal cell cancer. *Proceedings*
10 *of the National Academy of Sciences* **98**, 3387–3392 (2001).
- 11 9. Cha, P.-C., Takahashi, A., Hosono, N., Low, S.-K., Kamatani, N., Kubo, M. & Nakamura, Y. A
12 genome-wide association study identifies three loci associated with susceptibility to uterine
13 fibroids. *Nat. Genet.* **43**, 447–450 (2011).
- 14 10. Edwards, T. L., Michels, K. A., Hartmann, K. E. & Velez Edwards, D. R. BET1L and TNRC6B
15 associate with uterine fibroid risk among European Americans. *Hum. Genet.* **132**, 943–953
16 (2013).
- 17 11. Aissani, B., Zhang, K. & Wiener, H. Evaluation of GWAS candidate susceptibility loci for
18 uterine leiomyoma in the multi-ethnic NIEHS uterine fibroid study. *Front. Genet.* **6**, 241 (2015).
- 19 12. Bondagji, N. S., Morad, F. A., Al-Nefaei, A. A. A., Khan, I. A., Elango, R., Abdullah, L. S., M
20 Al-Mansouri, N., Sabir, J., Banaganapalli, B., Edris, S. & Shaik, N. A. Replication of GWAS loci
21 revealed the moderate effect of TNRC6B locus on susceptibility of Saudi women to develop
22 uterine leiomyomas. *J. Obstet. Gynaecol. Res.* **43**, 330–338 (2017).
- 23 13. Zhang, K., Wiener, H. & Aissani, B. Admixture mapping of genetic variants for uterine fibroids.
24 *J. Hum. Genet.* **60**, 533–538 (2015).
- 25 14. Eggert, S. L., Huyck, K. L., Somasundaram, P., Kavalla, R., Stewart, E. A., Lu, A. T., Painter, J.
26 N., Montgomery, G. W., Medland, S. E., Nyholt, D. R., Treloar, S. A., Zondervan, K. T., Heath,

- 1 A. C., Madden, P. A. F., Rose, L., Buring, J. E., Ridker, P. M., Chasman, D. I., Martin, N. G.,
2 Cantor, R. M. & Morton, C. C. Genome-wide linkage and association analyses implicate FASN
3 in predisposition to Uterine Leiomyomata. *Am. J. Hum. Genet.* **91**, 621–628 (2012).
- 4 15. Hellwege, J. N., Jeff, J. M., Wise, L. A., Gallagher, C. S., Wellons, M., Hartmann, K. E., Jones,
5 S. F., Torstenson, E. S., Dickinson, S., Ruiz-Narváez, E. A., Rohland, N., Allen, A., Reich, D.,
6 Tandon, A., Pasaniuc, B., Mancuso, N., Im, H. K., Hinds, D. A., Palmer, J. R., Rosenberg, L.,
7 Denny, J. C., Roden, D. M., Stewart, E. A., Morton, C. C., Kenny, E. E., Edwards, T. L. & Velez
8 Edwards, D. R. A multi-stage genome-wide association study of uterine fibroids in African
9 Americans. *Hum. Genet.* **136**, 1363–1373 (2017).
- 10 16. Makinen, N., Mehine, M., Tolvanen, J., Kaasinen, E., Li, Y., Lehtonen, H. J., Gentile, M., Yan, J.,
11 Enge, M., Taipale, M., Aavikko, M., Katainen, R., Virolainen, E., Bohling, T., Koski, T. A.,
12 Launonen, V., Sjoberg, J., Taipale, J., Vahteristo, P. & Aaltonen, L. A. MED12, the Mediator
13 Complex Subunit 12 Gene, Is Mutated at High Frequency in Uterine Leiomyomas. *Science* **334**,
14 252–255 (2011).
- 15 17. Heinonen, H.-R., Sarvilinna, N. S., Sjöberg, J., Kämpjärvi, K., Pitkänen, E., Vahteristo, P.,
16 Mäkinen, N. & Aaltonen, L. A. MED12 mutation frequency in unselected sporadic uterine
17 leiomyomas. *Fertil. Steril.* **102**, 1137–1142 (2014).
- 18 18. Mehine, M., Kaasinen, E., Heinonen, H.-R., Mäkinen, N., Kämpjärvi, K., Sarvilinna, N.,
19 Aavikko, M., Vähärautio, A., Pasanen, A., Bützow, R., Heikinheimo, O., Sjöberg, J., Pitkänen,
20 E., Vahteristo, P. & Aaltonen, L. A. Integrated data analysis reveals uterine leiomyoma subtypes
21 with distinct driver pathways and biomarkers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1315–1320
22 (2016).
- 23 19. Baird, D. M. Variation at the TERT locus and predisposition for cancer. *Expert Rev. Mol. Med.*
24 **12**, e16 (2010).
- 25 20. Abraham, G. & Inouye, M. Genomic risk prediction of complex human disease and its clinical
26 application. *Curr. Opin. Genet. Dev.* **33**, 10–16 (2015).

- 1 21. Rogalla, P., Rohen, C., Hennig, Y., Deichert, U., Bonk, U. & Bullerdiek, J. Telomere Repeat
2 Fragment Sizes Do Not Limit the Growth Potential of Uterine Leiomyomas. *Biochem. Biophys.*
3 *Res. Commun.* **211**, 175–182 (1995).
- 4 22. Bonatz, G., Frahm, S. O., Andreas, S., Heidorn, K., Jonat, W. & Parwaresch, R. Telomere
5 shortening in uterine leiomyomas. *Am. J. Obstet. Gynecol.* **179**, 591–596 (1998).
- 6 23. Edwards, T. L., Hartmann, K. E. & Velez Edwards, D. R. Variants in BET1L and TNRC6B
7 associate with increasing fibroid volume and fibroid type among European Americans. *Hum.*
8 *Genet.* **132**, 1361–1369 (2013).
- 9 24. Borahay, M. A., Al-Hendy, A., Kilic, G. S. & Boehning, D. Signaling Pathways in Leiomyoma:
10 Understanding Pathobiology and Implications for Therapy. *Mol. Med.* **21**, 242–256 (2015).
- 11 25. Biason-Lauber, A. & Konrad, D. WNT4 and sex development. *Sex Dev.* **2**, 210–218 (2008).
- 12 26. Hill, M. A. Embryology Uterus Development. (2018). at
13 <https://embryology.med.unsw.edu.au/embryology/index.php/Uterus_Development>
- 14 27. Yang, L., Han, Y., Suarez Saiz, F. & Minden, M. D. A tumor suppressor and oncogene: the WT1
15 story. *Leukemia* **21**, 1603–1603 (2007).
- 16 28. Markowski, D. N., Bartnitzke, S., Löning, T., Drieschner, N., Helmke, B. M. & Bullerdiek, J.
17 MED12 mutations in uterine fibroids--their relationship to cytogenetic subgroups. *Int. J. Cancer*
18 **131**, 1528–1536 (2012).
- 19 29. Vainio, S., Heikkilä, M., Kispert, A., Chin, N. & McMahon, A. P. Female development in
20 mammals is regulated by Wnt-4 signalling. *Nature* **397**, 405–409 (1999).
- 21 30. Peltoketo, H., Allinen, M., Vuosku, J., Kujala, S., Lundan, T., Salminen, A., Winqvist, R. &
22 Vainio, S. Characterization and expression of the human WNT4; lack of associated germline
23 mutations in high--to moderate--risk breast and ovarian cancer. *Cancer Lett.* **213**, 83–90 (2004).
- 24 31. Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L. M., Juodakis, J., Miller, D. E., Litterman,
25 N., Jiang, P.-P., Russell, L., Hinds, D. A., Hu, Y., Weirauch, M. T., Chen, X., Chavan, A. R.,
26 Wagner, G. P., Pavličev, M., Nnamani, M. C., Maziarz, J., Karjalainen, M. K., Rämetsä, M.,

- 1 Sengpiel, V., Geller, F., Boyd, H. A., Palotie, A., Momany, A., Bedell, B., Ryckman, K. K.,
2 Huusko, J. M., Forney, C. R., Kottyan, L. C., Hallman, M., Teramo, K., Nohr, E. A., Davey
3 Smith, G., Melbye, M., Jacobsson, B. & Muglia, L. J. Genetic Associations with Gestational
4 Duration and Spontaneous Preterm Birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).
- 5 32. Scelo, G., Purdue, M. P., Brown, K. M., Johansson, M., Wang, Z., Eckel-Passow, J. E., Ye, Y.,
6 Hofmann, J. N., Choi, J., Foll, M., Gaborieau, V., Machiela, M. J., Colli, L. M., Li, P., Sampson,
7 J. N., Abedi-Ardekani, B., Besse, C., Blanche, H., Boland, A., Burdette, L., Chabrier, A.,
8 Durand, G., Le Calvez-Kelm, F., Prokhortchouk, E., Robinot, N., Skryabin, K. G., Wozniak, M.
9 B., Yeager, M., Basta-Jovanovic, G., Dzamic, Z., Foretova, L., Holcatova, I., Janout, V., Mates,
10 D., Mukeriya, A., Rascu, S., Zaridze, D., Bencko, V., Cybulski, C., Fabianova, E., Jinga, V.,
11 Lissowska, J., Lubinski, J., Navratilova, M., Rudnai, P., Szeszenia-Dabrowska, N., Benhamou,
12 S., Cancel-Tassin, G., Cussenot, O., Baglietto, L., Boeing, H., Khaw, K.-T., Weiderpass, E.,
13 Ljungberg, B., Sitaram, R. T., Bruinsma, F., Jordan, S. J., Severi, G., Winship, I., Hveem, K.,
14 Vatten, L. J., Fletcher, T., Koppova, K., Larsson, S. C., Wolk, A., Banks, R. E., Selby, P. J.,
15 Easton, D. F., Pharoah, P., Andreotti, G., Freeman, L. E. B., Koutros, S., Albanes, D., Männistö,
16 S., Weinstein, S., Clark, P. E., Edwards, T. L., Lipworth, L., Gapstur, S. M., Stevens, V. L., Carol,
17 H., Freedman, M. L., Pomerantz, M. M., Cho, E., Kraft, P., Preston, M. A., Wilson, K. M.,
18 Michael Gaziano, J., Sesso, H. D., Black, A., Freedman, N. D., Huang, W.-Y., Anema, J. G.,
19 Kahnoski, R. J., Lane, B. R., Noyes, S. L., Petillo, D., Teh, B. T., Peters, U., White, E., Anderson,
20 G. L., Johnson, L., Luo, J., Buring, J., Lee, I.-M., Chow, W.-H., Moore, L. E., Wood, C., Eisen,
21 T., Henrion, M., Larkin, J., Barman, P., Leibovich, B. C., Choueiri, T. K., Mark Lathrop, G.,
22 Rothman, N., Deleuze, J.-F., McKay, J. D., Parker, A. S., Wu, X., Houlston, R. S., Brennan, P. &
23 Chanock, S. J. Genome-wide association study identifies multiple risk loci for renal cell
24 carcinoma. *Nat. Commun.* **8**, 15724 (2017).
- 25 33. Guleria, A. & Chandna, S. ATM kinase: Much more than a DNA damage responsive protein.
26 *DNA Repair* **39**, 1–20 (2016).

- 1 34. Lee, P. J., Yoo, N. S., Hagemann, I. S., Pfeifer, J. D., Cottrell, C. E., Abel, H. J. & Duncavage, E.
2 J. Spectrum of mutations in leiomyosarcomas identified by clinical targeted next-generation
3 sequencing. *Exp. Mol. Pathol.* **102**, 156–161 (2017).
- 4 35. Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat. Rev. Genet.* **6**,
5 611–622 (2005).
- 6 36. Liu, Y., Cao, L., Li, Z., Zhou, D., Liu, W., Shen, Q., Wu, Y., Zhang, D., Hu, X., Wang, T., Ye, J.,
7 Weng, X., Zhang, H., Zhang, D., Zhang, Z., Liu, F., He, L. & Shi, Y. A genome-wide association
8 study identifies a locus on TERT for mean telomere length in Han Chinese. *PLoS One* **9**, e85043
9 (2014).
- 10 37. Walsh, K. M., Codd, V., Smirnov, I. V., Rice, T., Decker, P. A., Hansen, H. M., Kollmeyer, T.,
11 Kosel, M. L., Molinaro, A. M., McCoy, L. S., Bracci, P. M., Cabriga, B. S., Pekmezci, M.,
12 Zheng, S., Wiemels, J. L., Pico, A. R., Tihan, T., Berger, M. S., Chang, S. M., Prados, M. D.,
13 Lachance, D. H., O'Neill, B. P., Sicotte, H., Eckel-Passow, J. E., ENGAGE Consortium Telomere
14 Group, van der Harst, P., Wiencke, J. K., Samani, N. J., Jenkins, R. B. & Wrensch, M. R. Variants
15 near TERT and TERC influencing telomere length are associated with high-grade glioma risk.
16 *Nat. Genet.* **46**, 731–735 (2014).
- 17 38. Lan, Q., Cawthon, R., Gao, Y., Hu, W., Dean Hosgood, H., Barone-Adesi, F., Ji, B.-T., Bassig,
18 B., Chow, W.-H., Shu, X., Cai, Q., Xiang, Y., Berndt, S., Kim, C., Chanock, S., Zheng, W. &
19 Rothman, N. Longer Telomere Length in Peripheral White Blood Cells Is Associated with Risk
20 of Lung Cancer and the rs2736100 (CLPTM1L-TERT) Polymorphism in a Prospective Cohort
21 Study among Women in China. *PLoS One* **8**, e59230 (2013).
- 22 39. Melin, B. S., Nordfjäll, K., Andersson, U. & Roos, G. hTERT cancer risk genotypes are
23 associated with telomere length. *Genet. Epidemiol.* **36**, 368–372 (2012).
- 24 40. Choi, B. J., Yoon, J. H., Kim, O., Choi, W. S., Nam, S. W., Lee, J. Y. & Park, W. S. Influence of
25 the hTERT rs2736100 polymorphism on telomere length in gastric cancer. *World J.*
26 *Gastroenterol.* **21**, 9328–9336 (2015).

- 1 41. Heinonen, H.-R., Pasanen, A., Heikinheimo, O., Tanskanen, T., Palin, K., Tolvanen, J.,
2 Vahteristo, P., Sjöberg, J., Pitkänen, E., Bützow, R., Mäkinen, N. & Aaltonen, L. A. Multiple
3 clinical characteristics separate MED12-mutation-positive and -negative uterine leiomyomas. *Sci.*
4 *Rep.* 7, 1015 (2017).

1 **TABLES**

2 **Table 1:** Predisposition loci for uterine leiomyoma.

Chr	Position	Band	rs-id	A	B	MAF	OR	P	Likely disease gene
11	108,143,456	q22.3	rs1800057	C	G	0.028	1.49	1.00E-16	<i>ATM</i>
11	32,364,187	p13	rs2057178	G	A	0.179	1.21	1.73E-13	<i>WT1</i>
6	152,558,197	q25.2	rs11751190	G	A	0.226	1.17	1.43E-11	<i>ESR1</i>
1	22,403,357	p36.12	rs12038474	G	A	0.202	1.17	1.81E-10	<i>WNT4</i>
22	40,652,873	q13.1	rs12484776 ^a	A	G	0.245	1.14	2.45E-08	?
5	1,286,516	p15.33	rs2736100 ^b	C	A	0.336	0.91	3.59E-07	<i>TERT</i>

74 The numbers for minor allele frequency (MAF), odds-ratio (OR) and association (P) are
75 based on the UKBB cohort and GBE query on self-reported and ICD identified UL cases
76 (n=5,417). Gene symbols are shown for reference; coordinates follow hg19. Genome-wide
77 significant P-values are shown bolded.

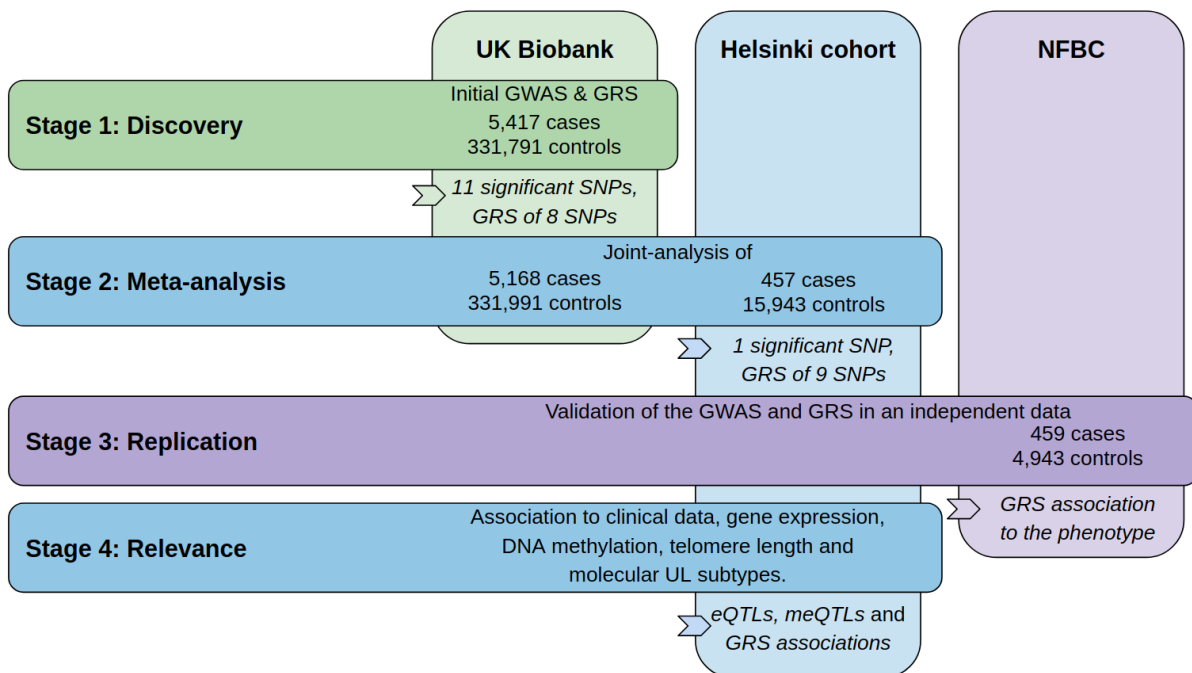
78 ^a Previously implicated predisposition to ULs.

79 ^b Previously associated to lung adenocarcinoma, glioma and telomere length.

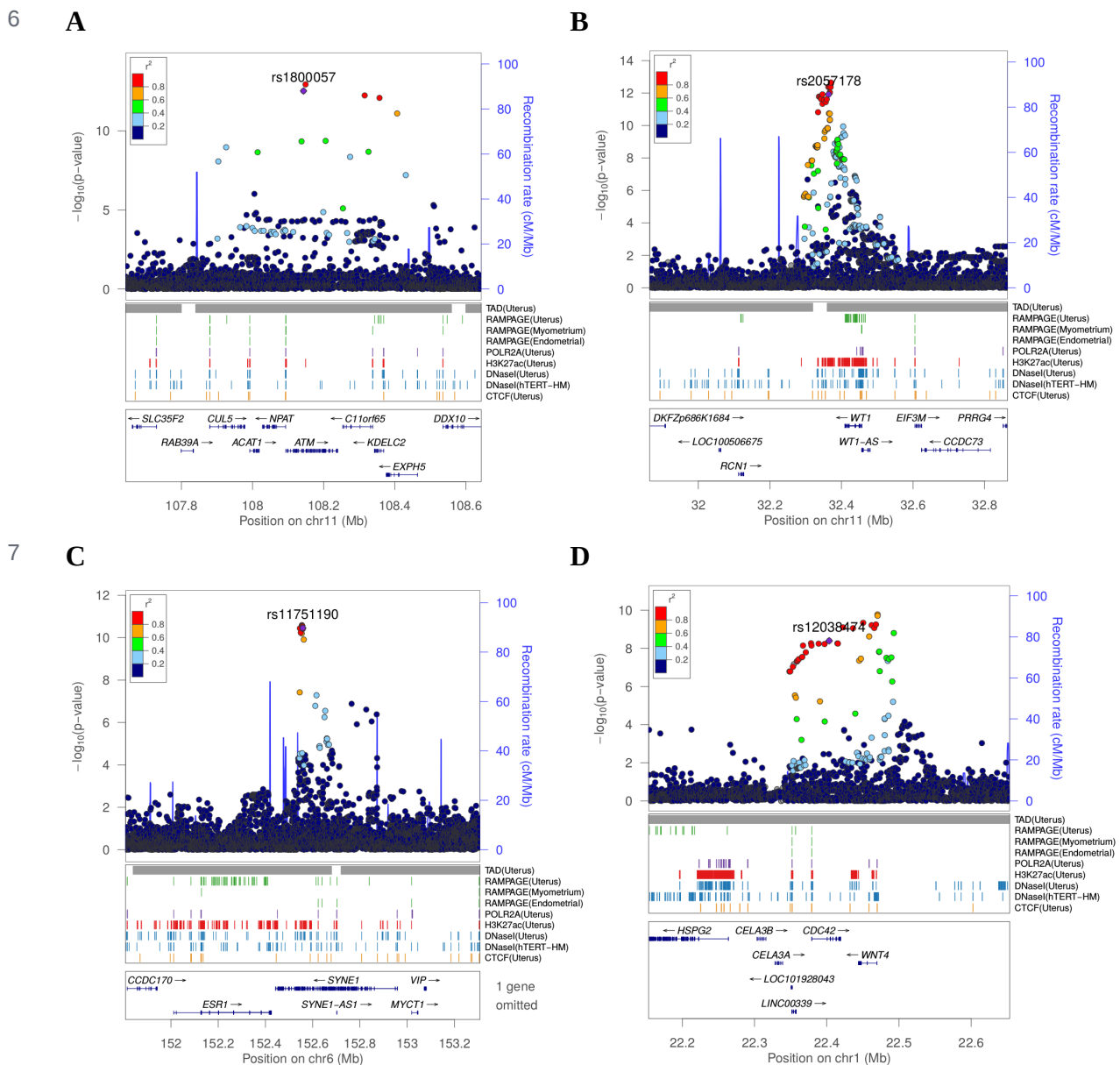
1 **FIGURES**

2 **Figure 1:** Outline of the study stages and genotyping cohorts. GRS, genomic risk score.

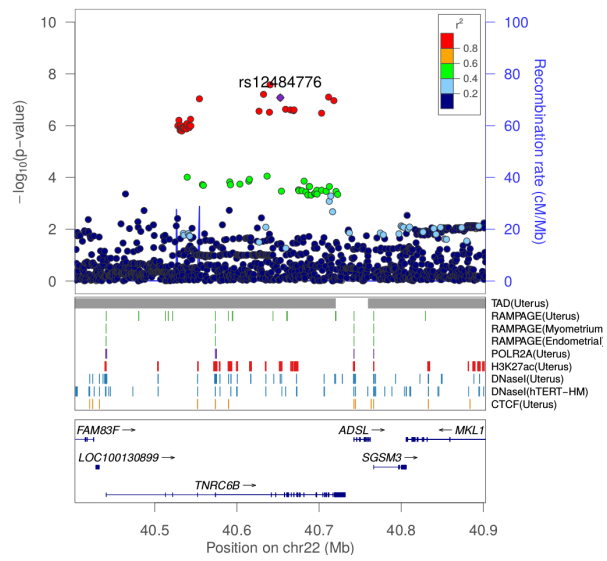
3 NFBC, Northern Finland Birth Cohort.



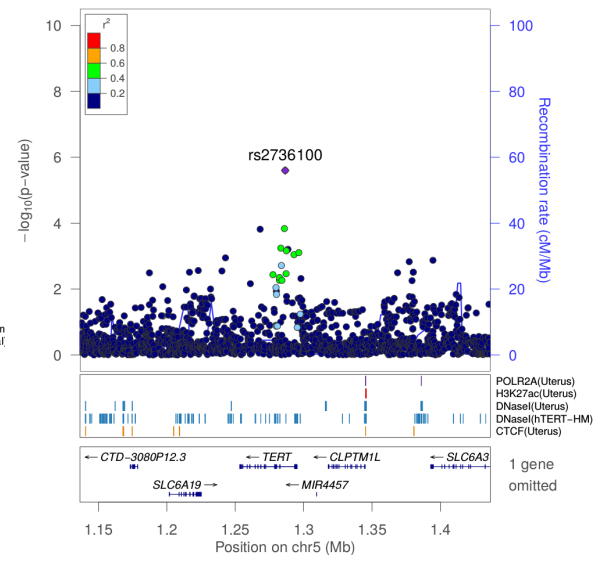
1 **Figure 2:** Structure of UL predisposition loci. **A-F**, the SNPs from stage 1 and their genomic
 2 context. The LD estimates (r^2) were taken from UK10k ALSPAC. Associations are based on
 3 the genome-wide summary statistics on self-reported UL cases ($n=5,168$; Neale lab data for
 4 10 million SNPs). Gene symbols and ENCODE tracks (details in Supplementary Methods)
 5 are shown for reference; coordinates follow hg19.



1 **E**



F



1 **Figure 3: Meta-analysis of UL risk revealed rs117245733 at a gene poor locus at 13q14.11.**

2 **A**, meta-analysis P-values and the genomic context at the locus. Gene symbols and ENCODE

3 tracks (details in Supplementary Methods) are shown for reference; coordinates follow hg19.

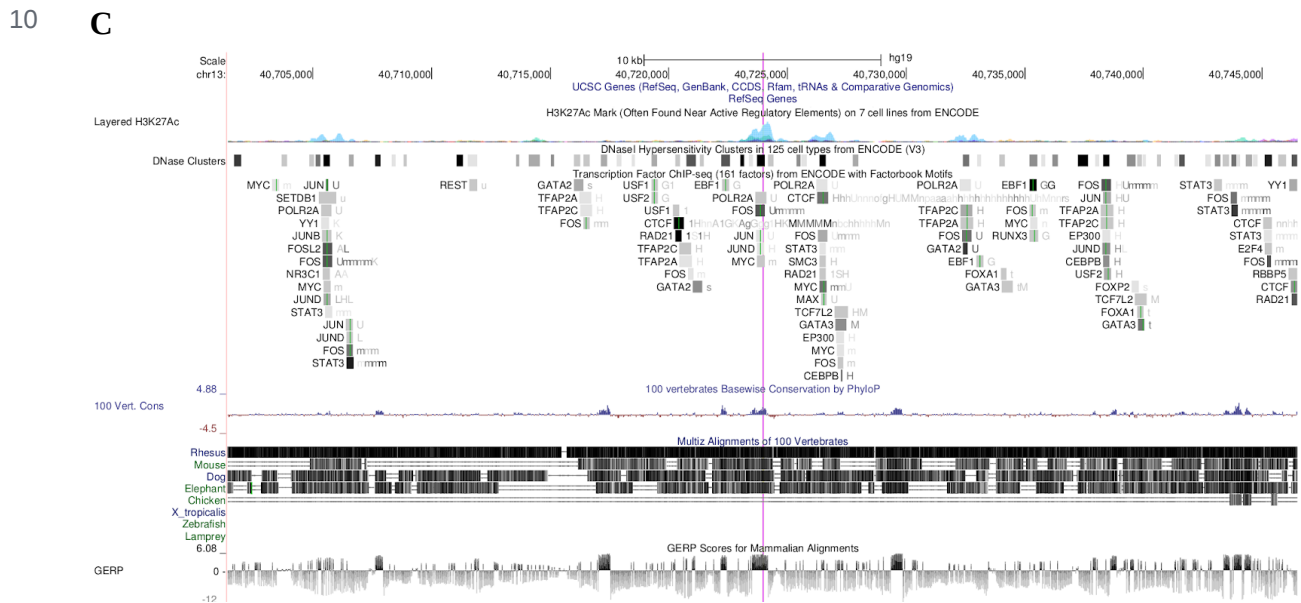
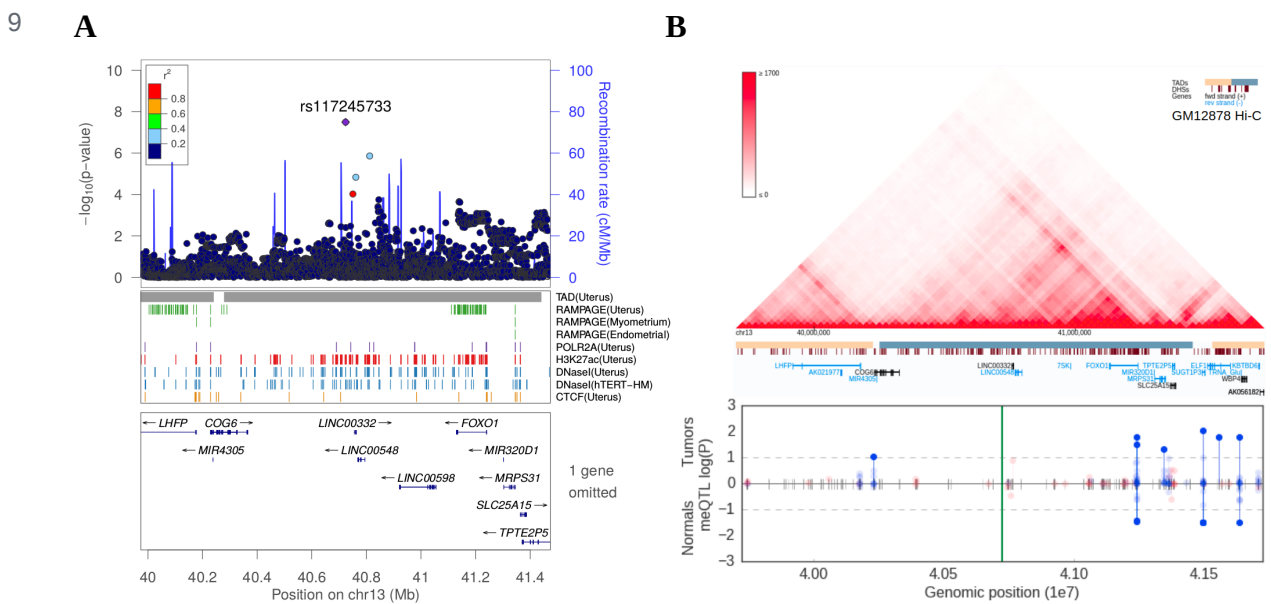
4 **B**, Hi-C, TADs and CpG methylation around the locus with an 1Mb flank. The needle plot

5 shows the meQTL associations (dashed lines at 10% FDR; green line denotes the SNP; gray

6 ticks denote all CpGs tested; blue needle for positive coefficient, red otherwise) for tumors

7 (above x-axis; $n_{AA}=53$, $n_{AB}=3$) and normals (below x-axis; $n_{AA}=33$, $n_{AB}=2$). **C**, UCSC genome

8 browser tracks related to conservation and regulation at the locus.



1 **Figure 4:** Methylation and expression differences in *WNT4*. **A**, Tumors (n=56) show
 2 methylation differences in three meQTLs (chr1:22456326, chr1:22456614 and
 3 chr1:22470407). **B**, Hi-C, TADs and CpG methylation around the locus with an 1Mb flank.
 4 The needle plot shows the meQTL associations (dashed lines at 10% FDR; green line denotes
 5 the two SNPs; gray ticks denote all CpGs tested; blue needle for positive coefficient,
 6 otherwise) for tumors (above x-axis; $n_{AA}=40$, $n_{AB}=15$, $n_{BB}=1$) and normals (below x-axis;
 7 $n_{AA}=23$, $n_{AB}=12$). **C**, *WNT4* locus. Hypomethylation refers to decreased methylation in BB vs.
 8 AA genotype and hypermethylation the opposite. The three CpGs from panel A are shown
 9 circled. **D**, Expression differences in tumors (n=41) *WNT4* based on rs61778046 genotype.

