Supplemental Methods:

Description for GWAS Custom Code: A Bayesian Sparse Linear Mixed Model (BSLMM) [37], available in the software program Genome-wide Efficient Mixed Model Association (GEMMA) [38], was used to rank SNPs based on the number of times in 10 million iterations that a SNP was estimated to have a large effect. This step was repeated 10 times and the maximum beta-value for each SNP was used for final ranking. Chromosomes were then divided into 500KB segments, and the top and two adjacent SNPs were kept within each segment. These SNPs were then used to build the select SNP GRM based on a stepwise feature selection, where each SNP was kept only if it is determined that it significantly improves the null model, which included both random and fixed effects. If a SNP was selected, inclusion of that SNP becomes the new null model for testing of the next SNP. Once the select SNP GRM was built, a linear mixed model, using the software program FaST-LMM [39], was performed with the select SNP GRM in place of the full GRM. FaST-LMM's algorithm tests each SNP individually for an effect on the phenotype using a maximum likelihood estimation [39]. SNPs within 1MB of the tested SNP were excluded from the select GRM to avoid double fitting of the model.

Description of LDAK software analyses: A full description of LDAK and the analyses available is available in Lee, *et. al.* [44,45]. Briefly, LDAK's algorithm uses restricted maximum likelihood to estimate the variance explained by all SNPs for a given phenotype. Unlike other mixed linear models available to estimate h2, LDAK incorporates an LD weighted genetic relationship matrix (LD-GRM) and a scaling factor to account for the effect of minor allele frequency on h2. The main output file includes

the h2 estimate, standard deviation (SD), log likelihood of the estimate, and the p-value for the log likelihood. For genomic partitioning, the LD-GRM is comprised only of SNPs from the specified region; h2 is then estimated from this subset of SNPs. For our analysis, we utilized our ROI identified on GWAS and *di* statistic, and included age, sex and section as covariates. The top predictors approach fit the top SNPs from the GWAS as covariates in the analysis using LDAK's --top-preds function. The output from this analysis includes the genetic variance explained by the top predictors, the genetic variance explained by the remainder of the SNPs, and then the sum of these values as the overall estimate of heritability. A SD is not provided for the top predictors.

Description of Random Sub-setting of Data: We also performed random sub-setting of the data by removing 10% of the population using the software package R's random number generator without replacement (R Core Team 2014). LD-GRMs were constructed from the reduced cohort and heritability estimates were calculated for total heritability, genomic partitioning of our ROI, and with the top ECA6 SNPs from the GWAS as covariates. This process was repeated 100 times and the average of all heritability estimates and SD were calculated and compared with the original estimates.

Description of TaqMan SNP Genotyping assay: A TaqMan SNP genotyping assay was utilized to efficiently genotype individuals for the HMGA2 c.83G>A variant. Forward (CTTCAGCCCAGGGACAAC) and reverse (AAGCAGCAGCAAGTCAGT) PCR primers were designed to produce an 80 base pair amplicon that included the HMGA c.83G>A variant. Locked nucleic acid (LNA) probes, with a 5' fluorescent reporter dye and 3'

quencher, were designed for allelic discrimination between the G allele (5HEX/AG+A+GA+G+G+ACG/3IABkFQ) and the A allele (56-FAM/AG+A+GA+G+A+A+CGC/3IABkFQ) as shown below.

GGCGGACTCGGGGCGGCTGAGGCCAGCGGCTGCAGCGGCGGTAGCGGCGGCG

GCGGCGGGAGGCAGGATGAGCGCACGCGGTGAGGGCGCCGGGCAGCCGTCCA**C**

TTCAGCCCAGGGACAACCTGCCGCCCCGGCGCCTCAGA**AGAGAGG/AACG**CGGC

CGACCCAGG**AAGCAGCAGCAAGTCAGT**ACGCGGGCGGGGTGGGGGCACCAGCC

CGCCTCCGCGCCCTCCGCGAAGGCCCGGCCACGCGCGGCCCCGAGCGCGGGAG

CCGGGTCGCCGCGCGCCGCCCGCCGGCCGGAGGCGGG

Supplemental Methods Figure 1:HMGA2 sequence approximately 150 bp 5' and 3' of the c.83G>A variant. The forward primer is indicated by green text and the reverse primer is indicated in red text. Orange text indicates sequence targeted by the florescent-tagged LNA probes.

Reaction components and volumes for each reaction were as follows: 5µL of 5ng/uL DNA template, 1µL forward primer, 1 µL reverse primer, 0.1 µL HEX probe, 0.1µL FAM probe, 5µL PrimeTime® Gene Expression Master Mix, 4µL betaine and 3µL molecular biology grade water for a final volume of 19.2µL. Final cycling protocol was as follows: Cycle 1 (1 repeat): 95^0C for 3 minutes. Cycle 2 (40 repeats): step 1: 95^0C for 5 seconds and step 2: 61^0C for 30 seconds.