Supplementary Online Content

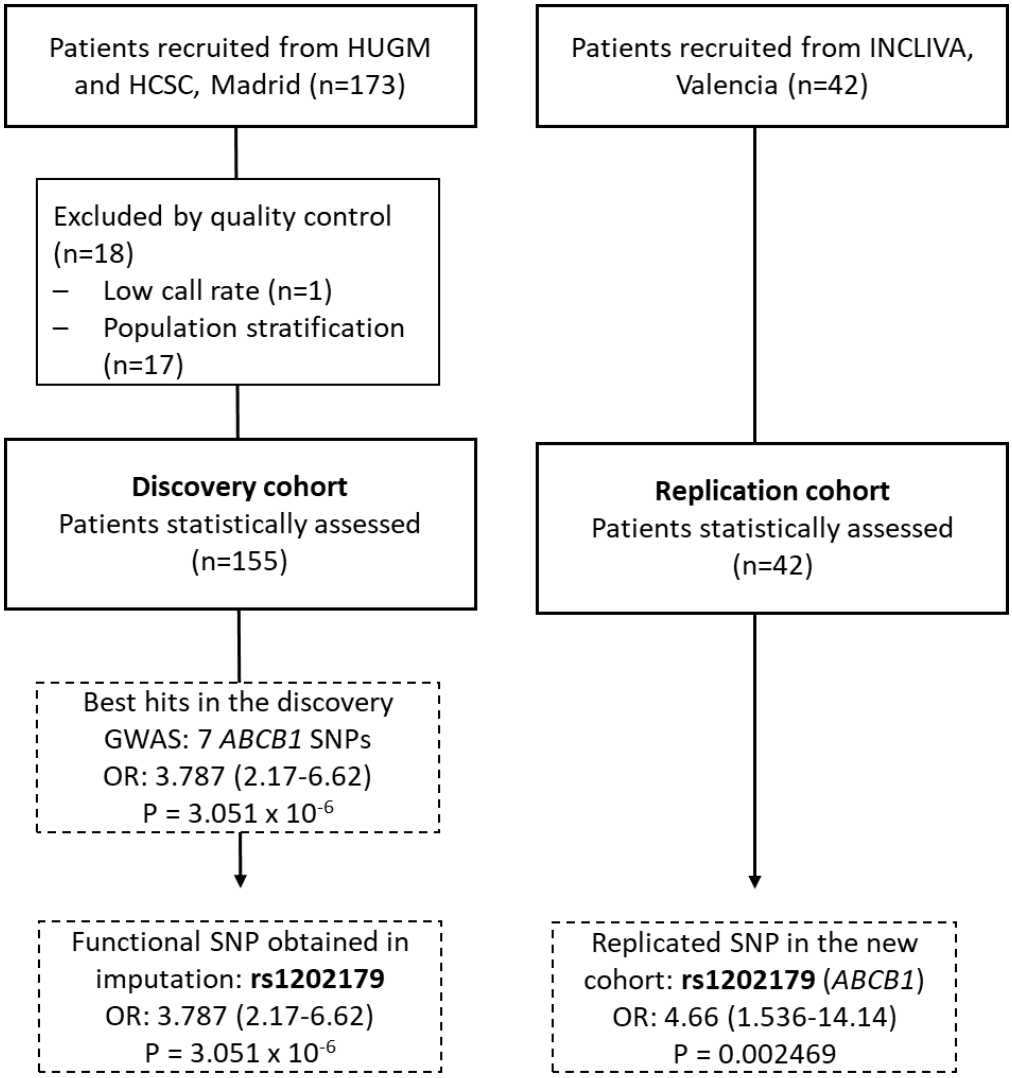
Núñez-Torres R, Martín M, García-Sáenz JÁ, et al. Association between *ABCB1* genetic variants and persistent chemotherapy-induced alopecia in women with breast cancer. *JAMA Dermatol*. Published online August 5, 2020. doi:10.1001/jamadermatol.2020.1867

This supplementary material has been provided by the authors to give readers additional information about their work.

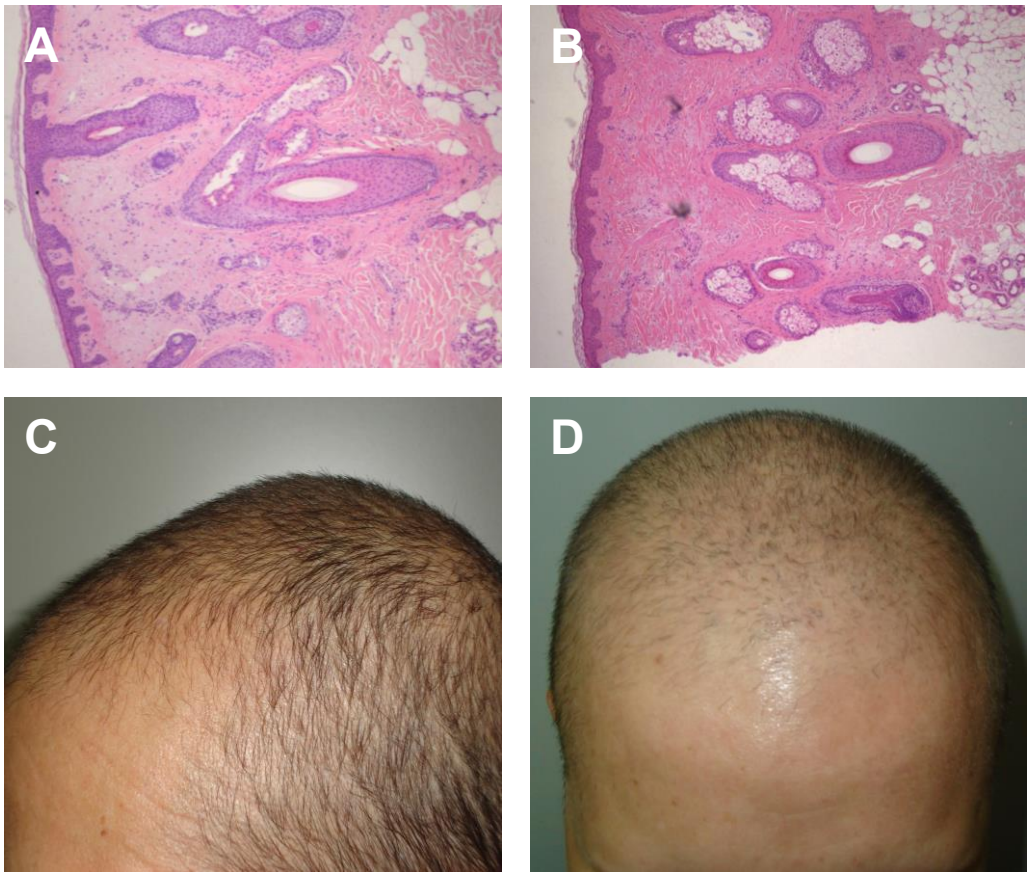
eAppendix 1. Study Population

Alopecia severity was graded using the Common Terminology Criteria for Adverse Events (CTCAE) v3.0, for acute (reversible) alopecia where: persistent chemotherapy-induced alopecia (pCIA) grade 1 is defined as weakening of hair or partial alopecia and not leading to the use of a wig after at least 18 months from the end of treatment; pCIA grade 2 is defined as complete alopecia that requires a wig 18 months after the end of treatment and grade 0 is defined as complete recovery of hair 12 months or more after the end of the treatment. No scalp-cooling device were applied in these patients. All controls received a cumulative docetaxel dose of >400mg/m^2. This inclusion criteria was applied because of previous evidence that no pCIA grade 2 cases were observed in patients treated with cumulative docetaxel doses below this threshold[1]. In total, 173 BC patients from Hospital Universitario Gregorio Marañón and Hospital Clínico San Carlos from Madrid (Spain) recruited between 2005 and 2014 were included in the discovery cohort (122 controls and 51 cases). For the replication phase, 42 BC patients (22 controls and 20 cases) recruited in 2018 from Hospital Clínico-INCLIVA from Valencia (Spain) were recruited (eFigure 1).

eFigure 1: Flowchart of the Patients Included in the Study



eFigure 2. A-B. Histological Images of Punch Biopsies Obtained From Persistent Chemotherapy-Induced Alopecia (pCIA) Grade 2 Patients. C-D. Clinical Photos of Grade 2 pCIA Patients Scalp.



eAppendix 2. Genome-Wide Association Study

2.1. Methods

2.1.1. Genotyping

DNA samples were quantified using Picogreen method (Invitrogen). A total amount of 250 ng of DNA was processed according to the Infinium HTS assay Protocol (Part # 15045738 Rev. A, Illumina), including amplification, fragmentation and hybridization using the HumanOmniExpress BeadChip. The array contains a total of 713,599 markers and was scanned on an iScan platform (Illumina).

2.1.2. Quality Control

Genotyping quality control was conducted with PLINK v1.9[2]. Samples with a call rate <0.95, SNPs with a genotyping rate <0.95 and SNPs whose genotype frequencies in control individuals departed from Hardy-Weinberg equilibrium at $P<10^{-6}$ were excluded from further analysis. Non-european ancestry individuals were identified using a principal component analysis (PCA) including samples from 1000 Genomes project (http://www.internationalgenome.org/) from European (CEU), Iberian (IBS), Southern Han Chinese (CHS) and Yoruba (YRI) populations using R statistical environment version 3.3.0 (www.r-project.org) and subsequently exclude from the analysis. A quantile-quantile (Q-Q) plot comparing the distribution of P-values observed to those expected was generated using the qqman package in R (v.3.3.0), and the genomic inflation factor (λ) derived from the Q-Q plot was calculated using PLINK v1.9[2].

2.1.3. Statistical Analysis

Quantitative variables in discovery and validation samples were compared between cases and controls by means of Student's *t*-test or the Mann–Whitney *U* test (for normally and not normally distributed data, respectively). Comparisons between categorical variables were performed using Fisher's exact test. All analyses were two tailed and p-values<0.05 were considered significant. Statistical analysis of clinical data was carried out with SPSS 19.0.

Logistic regression was performed using PLINK v1.9 without covariates and with those factors associated with pCIA in the univariate analysis with a *P*-value ≤ 0.05 to control the possible influence of cofounder factors in the association using step-wise logistic regression in SPSS 19.0.

A threshold of $5x10^{-6}$ was established to determine the associated markers for further replication in the validation cohort.

2.1.4. SNP Imputation

For the significant loci identified in the discovery phase, we performed predictions of the untyped SNPs within the LD block. Data were phased with ShapeIT[3] and imputed using Impute2[4] with 1000 Genomes Phase 3 build 37 (release date October 2014) as the reference. SNPs with Impute2 info metric >0.3 and imputed genotypes with uncertainty less than 0.1 were treated as hard calls and therefore tested for association by logistic analysis.

Detailed inspection of the region of interest was carried out with LocusZoom (http://locuszoom.org/) showing 50kb surrounding the SNP of interest.

2.1.5. Functional annotation

Functional annotations for strongly associated SNPs were explored in the following public databases: GTEX (gtexportal.org) and Phenoscanner (www.phenoscanner.medschl.cam.ac.uk) to evaluate the impact of the SNP on gene expression; LDlink[5] to obtain linkage disequilibrium data; Haploreg (pubs.broadinstitute.org/mammals/haploreg/haploreg.php) for exploring haplotype block as well as regulatory data; UCSC genome browser (https://genome.ucsc.edu/) to explore the genomic landscape and to integrate all genomic and functional data, and GeneHancer[6] to obtain predictions of enhancer interactions.

The boundaries for the linkage disequilibrium (LD) block containing the most significant signals were defined by the furthest upstream and downstream SNPs that correlated with the best hit (r^2>0.80) based on data from Phase 3 of the 1000 Genomes Project (http://phase3browser.1000genomes.org). LD information presented is based on the European population.

2.1.6. Haplotype analysis

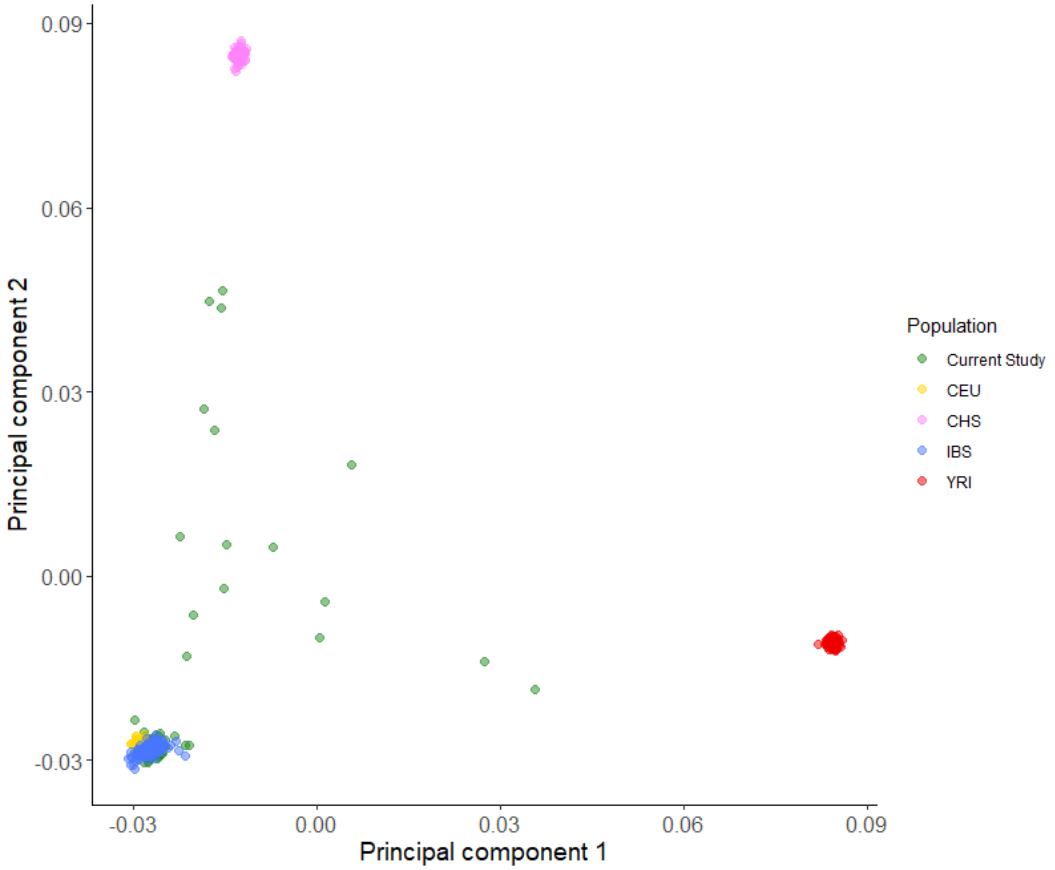
Genotypes were extracted from PLINK and haplotypes were estimated using PHASE v2.1.1. The risk haplotype according to the literature was evaluated, and haplotype frequencies in cases and controls as well as the association of the risk haplotype with pCIA were evaluated by logistic regression using SPSS 19.

2.2. Results

2.2.1 GWAS results

After quality control analysis in the discovery cohort, only one out of 173 samples was excluded due to low call rate (<95%) and 17 samples were excluded due to non-European ancestry origin showed in the PCA (eFigure 3), leaving 155 patients (47 cases and 108 controls) for further analyses.

eFigure 3. Principal Component Analysis of Discovery Cohort



Principal component analysis (PCA) of genetic data for the 173 samples from the discovery cohort (green dots) and SNPs that passed quality control (693,898 SNPs) including CEU (yellow dots), IBS (blue dots), CHS (pink dots) and YRI (red dots) populations from the 1000 Genomes project (http://www.internationalgenome.org/). Principal component 1 and principal component 2 were plotted.

Clinical characteristics of the 155 patients finally included in the discovery cohort after quality control analysis were not different from the initial cohort (eTable 1).

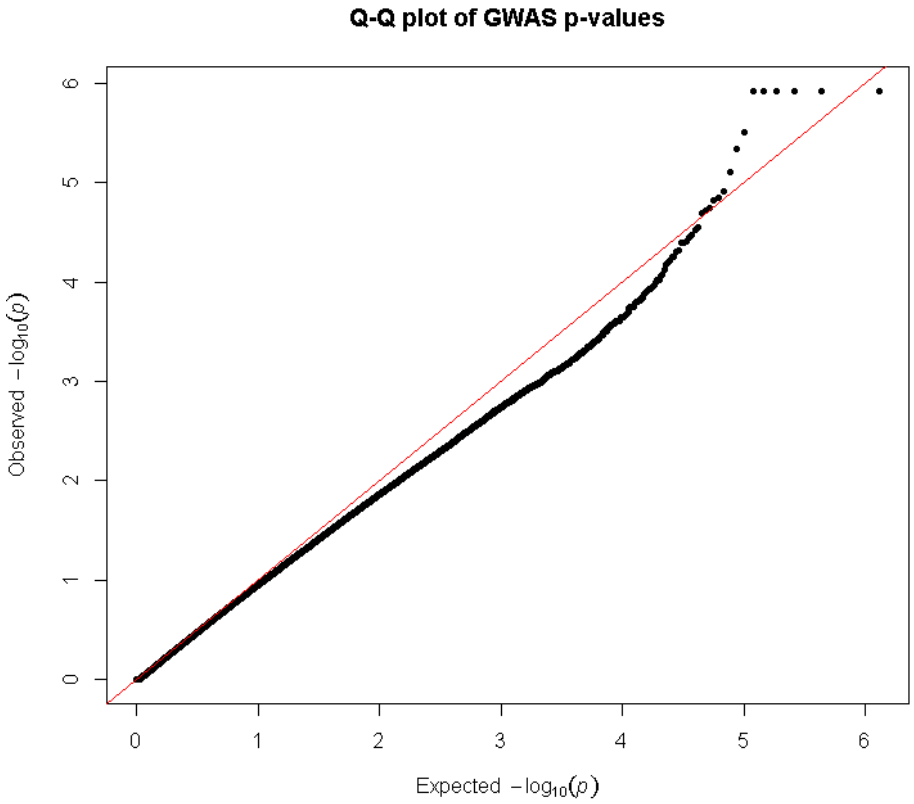
eTable 1. Summary of Main Characteristics of Patients Included in the GWAS Analysis After Quality Control Analysis

Characteristic	Cases (n = 47)	Controls (n = 108)	*P* Value
Age at time of treatment, median (IQR)[a]	57 (46-62)	52 (45-62)	0.003
Cumulative dose, median (range), mg/m^2	433.9 (400-450)	423.6 (400-450)	0.004
Aromatase inhibitors (AIs), No (%)[a]	60 (49)	63 (51)	0.568
Anthracyclines treatment, No (%)	15 (31.9)	75 (69.4)	1.34×10^{-5}

Abbreviations: IQR, interquartile range.

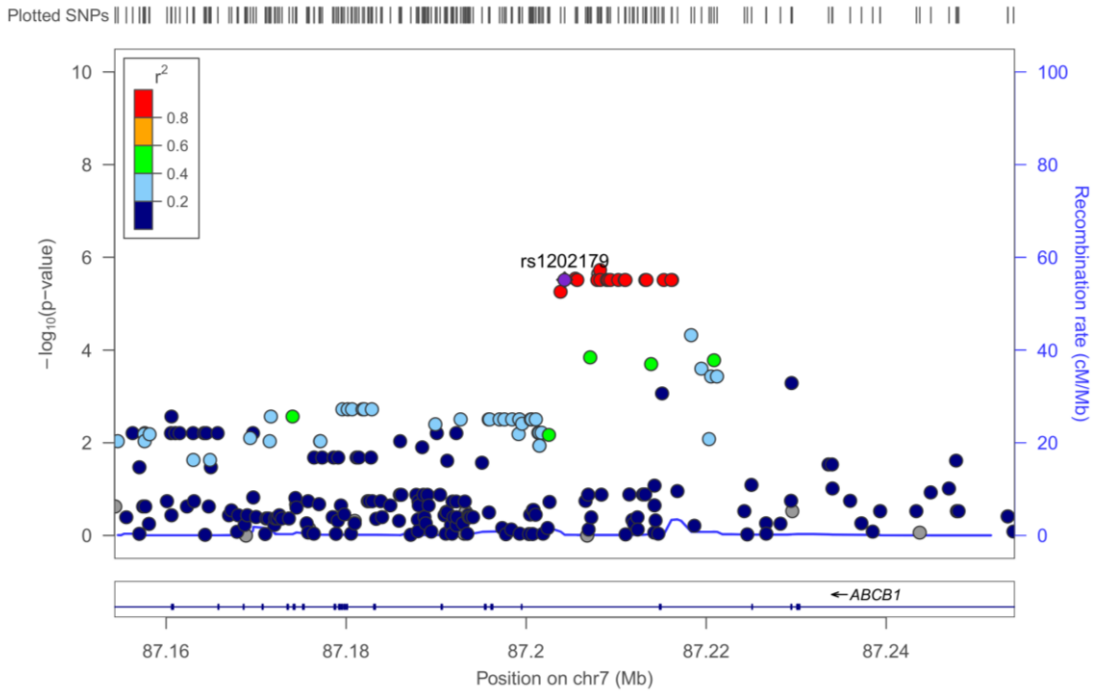
[a]For patients with available data.

efigure 4. Quantile-Quantile Plot of the Discovery Cohort



Q-Q plot showing the distributions of observed −log10(P-values) from logistic regression analysis for 693,898 SNPs in the discovery cohort plotted against expected −log10(P-values). Smaller P-values than would be expected by chance are observed at the tail of the plot. Plink-estimated genomic inflation factor λ= 1 indicates no obvious population stratification.

eFigure 5. Association and Recombination Plot of the Risk Locus and ± 50-kb Boundaries



Plot shows the genomic region associated with pCIA and the −log10 (P-values) of genotyped and imputed SNPs. Circles indicate genotyped and imputed SNPs and the diamond indicates the selected SNP, rs1202179. Recombination rates are also shown. SNP color indicates the strength of LD (r^2) with rs1202179. Recombination rates are based on the 1000 Genomes Project and genomic coordinates are based on Genome Reference Consortium Human Build 37 (GRCh37/hg19).

2.2.2. Statistical analysis with clinical variants

In addition to age and docetaxel cumulative dose the use of anthracyclines was associated with pCIA phenotype in the univariate analysis (*p-value*=5.79x10^{-7}). For this reason, a multivariate step-wise logistic regression was performed to estimate the association between the risk of pCIA and rs1202179 adjusting for the associated cofounding factors identified in the univariate analysis. However, the step-wise analysis exclude docetaxel cumulated doses as covariate in the model since it is correlated with the use of anthracyclines (*Spearman* ρ=-0.726 *p-value*=1.25x10^{-26}), including in the final model only age and use of anthracyclines therapy (OR: 5.28; 95%CI: 2.65-10.54; P=2.262 x 10^{-6}). This correlation could be explained due to the fact that the vast majority of the patients treated with anthracyclines (95.4%) received low cumulated docetaxel doses (CD<450 mg/m^2) (*P-value*=5.56x10^{-32}).

2.2.3. Fine-mapping analysis

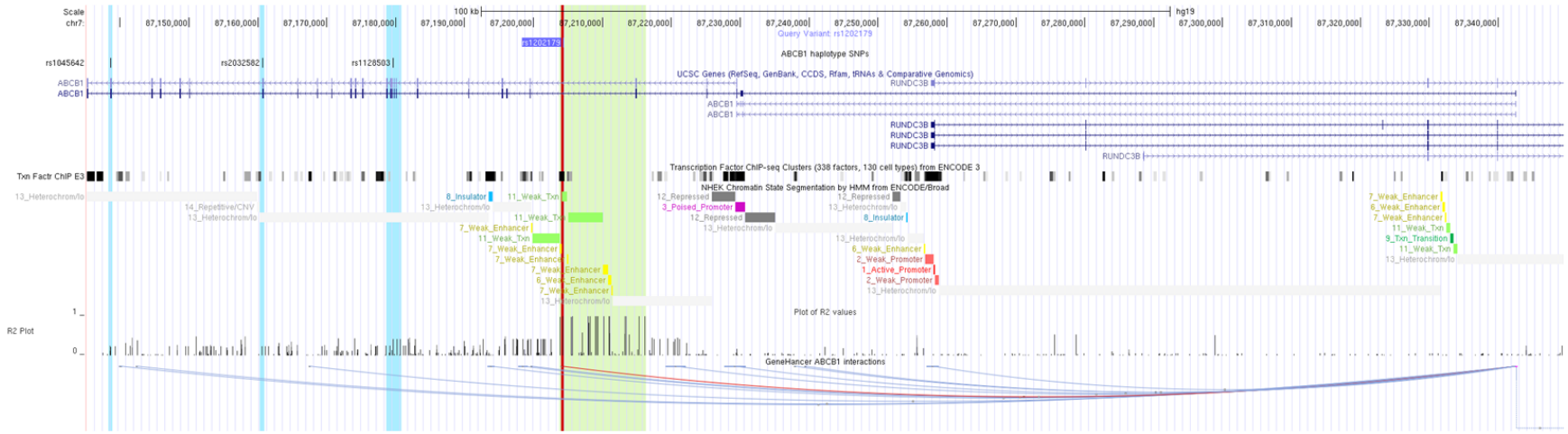
To fine-map the most significant signal at the 7q21.1 locus, the whole chromosome 7 was imputed resulting in the assignment of 4,741,475 SNPs. After quality control analysis, 977,567 SNPs were reliably imputed. We found 12 imputed SNPs strongly correlated with genotyped SNPs to be associated with pCIA risk (Table2) within the LD block of 11 kb spanning positions 87203840-87216216 in chromosome 7 (GRCh37/hg19).

Potential functional roles for these imputed ABCB1 SNPs and the 7 ABCB1 genotyped SNPs in the LD block were explored in order to identify the most likely causal variant driving pCIA risk. The functional annotations using data from ENCODE are depicted in eFigure6. We found that the imputed SNP rs1202179 is located in a regulatory region which, according to GeneHancer data[6], interacts with the *ABCB1* promoter (eFigure6), suggesting a possible regulatory role in ABCB1 expression. In addition, we found a correlation between the rs1202179 genotype and ABCB1 expression based on data of 31,640 individuals (P = 1.64 x10^{-20})[7]. In particular, rs1202179 risk allele C was associated with a decreased ABCB1 mRNA expression.

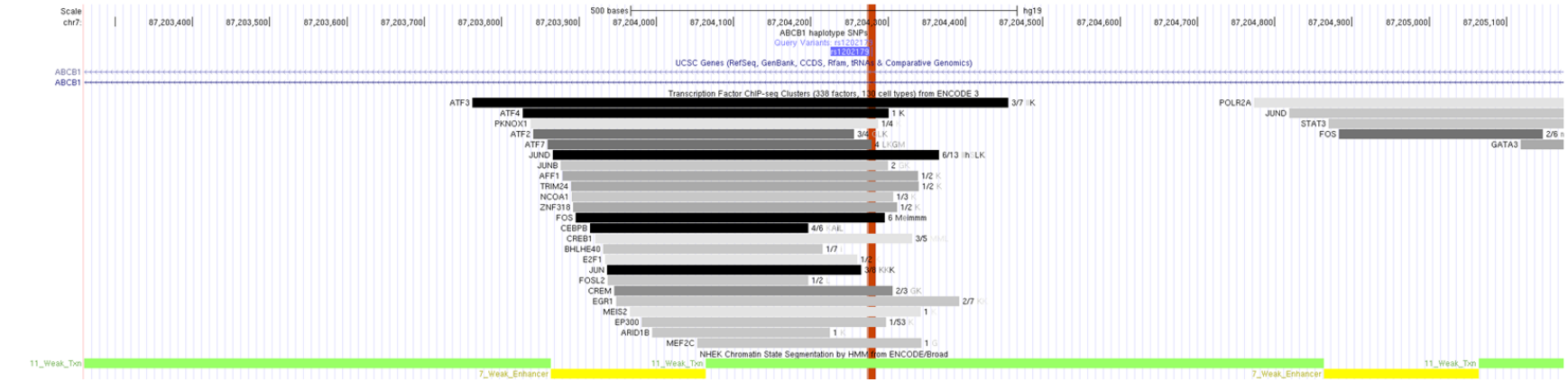
After selection of rs1202179, imputed genotypes were corroborated using allelic discrimation (See eAppendix 3 Section 3.1). All the genotypes were validated excepting 1 patient who shows rs1202179-TC genotype instead of the rs1202179-TT genotype imputed. All the logistic regression performed were corrected including the real genotype instead of the imputed one.

eFigure 6. Functional Annotation of the 7q21.1 Risk Locus Using Data From ENCODE

A



B



A) UCSC tracks with functional information of ABCB1. SNP rs1202179 is highlighted in red, the ABCB1 LD block containing SNPs in high LD (R2>0.97) with rs1202179 is highlighted in green, three well-known ABCB1 SNPs are highlighted in blue. Tracks showing (from top to bottom): chromosomal position (GRCh37/hg19); position of the three well-known ABCB1 SNPs; position of the ABCB1 SNPs identified in this study; genomic landscape of ABCB1 and RUNDC3B genes; transcription factor domains reported by ChIP-seq from ENCODE 3; Chromatin State in the NHEK (Normal Human Epidermal Keratinocytes) cell line from ENCODE; R2 values refer to rs1202179; predicted interactions with ABCB1 promoter from GeneHancer. B) Zoom to region surrounding genomic location of rs1202179 (highlighted in red). rs1202179 is located in several transcription factor binding sites.

2.2.4. Well-known ABCB1 variants and haplotype analysis

The most common variants studied –rs1128503 (1236T>C, Gly412Gly), rs2032582 (2677T>G/A, Ser893Ala/Thr) and rs1045642 (3435T>C, Ile1145Ile)[8]– were explored, and none of them reached the threshold of significance of $5x10^{-6}$ established in this study: rs1128503 (OR: 0.4024; 95%CI: 0.2266- 0.7144; P = 0.001883); rs2032582 (OR: 0.4235; 95%CI: 0.2416- 0.7423; P = 0.002692); and rs1045642 (OR: 0.6726; 95%CI: 0.4031- 1.122; P = 0.1288). These significances completely disappear when rs1202179 is included into the model as a covariate: rs1128503 (OR: 0.8988; 95%CI: 0.4228- 1.91; P = 0.7815); rs2032582 (OR: 1.047; 95%CI: 0.4861- 2.254; P = 0.9072); and rs1045642 (OR: 1.194; 95%CI: 0.6425- 2.219; P = 0.575).

In addition, we explored the association with the well-known ABCB1 haplotype including these three SNPs (rs1128503, rs2032582 and rs1045642). The haplotype defined by rs1128503-C, rs2032582-G and rs1045642-C was identified as the most frequent in our population and was therefore selected as the reference haplotype (eTable2). A significant association between the rs1045642-T-rs2032582-T-rs1128503-T haplotype and pCIA was found (OR: 042; 95%CI: 0.23 - 0.77; P=0.0049) (eTable2). However, this significant association could be explained by the correlation between the variants belonging to these haplotypes and the strongest associated variant rs1202179 (rs1045642 r^2=0.158504, rs2032582 r^2=0.396021 and rs1128503 r^2= 0.362534)

eTable 2. Well-Known *ABCB1* Haplotype Frequencies and Their Association With pCIA

Haplotype	Controls (n=108)	Cases (n=98)	OR (95% CI)	P-value
C-G-C (Reference)	100 (0.46)	56 (0.6)		
T-T-T	77 (0.36)	18 (0.19)	0.42 (0.23-0.77)	0.0049
T-G-C	21 (0.10)	16 (0.17)	1.36 (0.66-2.82)	0.4072
C-T-C	9 (0.04)	3 (0.03)	0.6 (0.15-2.29)	0.4503
C-G-T	6 (0.03)	1 (0.01)	0.3 (0.03-2.53)	0.2675
C-T-T	3 (0.01)	0	NA	NA

Abbreviations: OR, Per-allele odds ratio associated with effect allele; 95%CI, 95% Confidence Interval; NA, not available. Haplotype was defined by the following SNPs: rs1045642-rs2032582-rs1128503

eAppendix 3. Replication Analysis

3.1. Methods

Allelic discrimination assays were used to validate the imputed genotypes for rs1202179 and to genotype the variant in the validation cohort. We used TaqMan SNP Genotyping Assay C___2982713_10 (ThermoFisher Scientific) using the manufacturer's recommended standard conditions. Fluorescence was detected in the 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific). Sequence Detection Systems (SDS) software v2.4 was used to analyze the data and the genotype assignation.

Logistic regression analysis considering age and use of anthracyclines as covariates was performed in the replication cohort using PLINK v1.9[2].

3.2. Results

Logistic regression analysis in replication cohort including age and anthracyclines-combined therapy as covariates, according to the model used in the discovery phase (See eAppendix 2, Section 2.2.2. for details), shows similar results (OR: 5.247; 95%CI: 1.581- 17.42; P= 0.006762). At the end, we performed the logistic analysis with the clinical covariates in the whole cohort showing the association of the rs1202179 with pCIA phenotype (OR: 4.455; 95%CI: 2.594- 7.65; P= 6.151×10^{-08}).

eReferences

1. Martín M, de la Torre-Montero JC, López-Tarruella S, et al. Persistent major alopecia following adjuvant docetaxel for breast cancer: incidence, characteristics, and prevention with scalp cooling. *Breast Cancer Res Treat*. 2018;171(3):627-634. doi:10.1007/s10549-018-4855-2

2. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

3. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9(2):179-181. doi:10.1038/nmeth.1785

4. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6). doi:10.1371/journal.pgen.1000529

5. J. Machiela M, J. Chanock S. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-3557. doi:10.1093/bioinformatics/btv402

6. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017;2017:1-17. doi:10.1093/database/bax028

7. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv*. 2018;18:10. doi:10.1101/447367

8. Wolking S, Schaeffeler E, Lerche H, Schwab M, Nies AT. Impact of Genetic Polymorphisms of ABCB1 (MDR1, P-Glycoprotein) on Drug Disposition and Potential Clinical Implications: Update of the Literature. *Clin Pharmacokinet*. 2015;54(7):709-735. doi:10.1007/s40262-015-0267-1