Genome-wide association study identifies novel susceptibility loci for *KIT* D816V positive mastocytosis

Gabriella Galatà,[1] Andrés C. García-Montero,[2,3] Thomas Kristensen,[4,5] Ahmed A.Z. Dawoud,[1] Javier I. Muñoz-González,[2,3] Manja Meggendorfer,[6] Paola Guglielmelli,[7] Yvette Hoade,[1] Ivan Alvarez-Twose,[8] Christian Gieger,[9,10,11,12] Konstantin Strauch,[9,13,14] Luigi Ferrucci,[15] Toshiko Tanaka,[15] Stefania Bandinelli,[16] Theresia M. Schnurr,[17] Torsten Haferlach,[6] Sigurd Broesby-Olsen,[5,18,19] Hanne Vestergaard,[5,20] Michael Boe Møller,[4,5] Carsten Bindslev-Jensen,[5,18,19] Alessandro M. Vannucchi,[7] Alberto Orfao,[2,3] Deepti Radia,[21] Andreas Reiter,[22] Andrew J. Chase,[1,23] Nicholas C.P. Cross,[1,23,24,*] and William J. Tapper[1,24]

Summary

Mastocytosis is a rare myeloid neoplasm characterized by uncontrolled expansion of mast cells, driven in >80% of affected individuals by acquisition of the *KIT* D816V mutation. To explore the hypothesis that inherited variation predisposes to mastocytosis, we performed a two-stage genome-wide association study, analyzing 1,035 individuals with *KIT* D816V positive disease and 17,960 healthy control individuals from five European populations. After quality control, we tested 592,007 SNPs at stage 1 and 75 SNPs at stage 2 for association by using logistic regression and performed a fixed effects meta-analysis to combine evidence across the two stages. From the meta-analysis, we identified three intergenic SNPs associated with mastocytosis that achieved genome-wide significance without heterogeneity between cohorts: rs4616402 ($p_{meta} = 1.37 \times 10^{-15}$, OR = 1.52), rs4662380 ($p_{meta} = 2.11 \times 10^{-12}$, OR = 1.46), and rs13077541 ($p_{meta} = 2.10 \times 10^{-9}$, OR = 1.33). Expression quantitative trait analyses demonstrated that rs4616402 is associated with the expression of *CEBPA* ($p_{eQTL} = 2.3 \times 10^{-14}$), a gene encoding a transcription factor known to play a critical role in myelopoiesis. The role of the other two SNPs is less clear: rs4662380 is associated with expression of the long non-coding RNA gene *TEX41* ($p_{eQTL} = 2.55 \times 10^{-11}$), whereas rs13077541 is associated with the expression of *TBL1XR1,* which encodes transducin (β)-like 1 X-linked receptor 1 ($p_{eQTL} = 5.70 \times 10^{-8}$). In individuals with available data and non-advanced disease, rs4616402 was associated with age at presentation ($p = 0.009$; beta = 4.41; n = 422). Additional focused analysis identified suggestive associations between mastocytosis and genetic variation at *TERT*, *TPSAB1/TPSB2*, and *IL13*. These findings demonstrate that multiple germline variants predispose to *KIT* D816V positive mastocytosis and provide novel avenues for functional investigation.

Introduction

Mastocytosis (MIM: 154800) is an uncommon myeloid neoplasm characterized by expansion and accumulation of clonal mast cells in one or more organ systems, including bone marrow, skin, liver, spleen, and gastrointestinal tract. The extent of organ infiltration and organ damage serves as the basis for classification as cutaneous mastocytosis (CM) or systemic mastocytosis (SM).[1] CM is typically found in children, while most adults with mastocytosis have SM with involvement of the bone marrow. Six main subtypes of SM are recognized: indolent SM (ISM) and smoldering systemic mastocytosis (SMM) are relatively benign forms that usually have a stable clinical course over many years. In contrast, SM with an associated hematologic neoplasm (SM-AHN), aggressive SM (ASM), and mast cell

[1]School of Medicine, University of Southampton, Southampton SO17 1BJ, UK; [2]Institute of Biomedical Research of Salamanca, Salamanca 37007, Spain; [3]Servicio de Citometría, Departamento de Medicina, CIBERONC, and Instituto de Biología Molecular y Celular del Cáncer, CSIC/Universidad de Salamanca, Salamanca 37007, Spain; [4]Department of Pathology, Odense University Hospital, 5000 Odense, Denmark; [5]Mastocytosis Centre Odense University Hospital, 5000 Odense, Denmark; [6]Munich Leukemia Laboratory, 81377 Munich, Germany; [7]Centro di Ricerca e Innovazione per le Malattie Mieloproliferative, Azienda Ospedaliera Universitaria Careggi, Dipartimento di Medicina Sperimentale e Clinica, Università Degli Studi di Firenze, 50134 Firenze, Italy; [8]Instituto de Mastocitosis de Castilla La Mancha, Hospital Virgen del Valle, 45071 Toledo, Spain; [9]Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany; [10]German Centre for Cardiovascular Research Partner Site Munich Heart Alliance, 80802 Munich, Germany; [11]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Germany Research Center for Environmental Health, 85764 Neuherberg, Germany; [12]German Center for Diabetes Research, 85764 Neuherberg, Germany; [13]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, 80539 Munich, Germany; [14]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg University, 55131 Mainz, Germany; [15]Longitudinal study section, Translation Gerontology Branch, National Institute on Aging, Baltimore, MD 21224, USA; [16]Geriatric Unit, Azienda USL Toscana centro, 50137 Firenze, Italy; [17]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark; [18]Department of Dermatology and Allergy Centre, Odense University Hospital, 5000 Odense, Denmark; [19]Odense Research Center for Anaphylaxis, Odense University Hospital, 5000 Odense, Denmark; [20]Department of Hematology, Odense University Hospital, 5000 Odense, Denmark; [21]Department of Clinical Haematology, Guy's and St Thomas' NHS Hospitals, London SE1 9RT, UK; [22]University Hospital Mannheim, Heidelberg University, 68167 Mannheim, Germany; [23]Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury SP2 8BJ, UK
[24]These authors contributed equally
*Correspondence: ncpc@soton.ac.uk
https://doi.org/10.1016/j.ajhg.2020.12.007.

leukemia (MCL), collectively known as advanced SM (advSM), are associated with a poor prognosis.[2] ISM is the most common of the six subtypes, accounting for 80% of SM-affected individuals.[3]

Approximately 80%–90% of adult SM-affected individuals across all subtypes test positive for the somatic mutation *KIT* c.2447A>T (p.Asp816Val), usually referred to as *KIT* D816V. Due to the nature of the disease, the mutant allele frequency is often very low, particularly in peripheral blood samples, and sensitive methods are needed for its detection.[4] *KIT* D816V mutation burden, serum tryptase, and β2-microglobulin levels correlate with disease burden and severity,[5–8] and for advSM, additional somatic mutations in *SRSF2*, *ASXL1*, and *RUNX1* indicate an adverse prognosis.[9–11]

Mastocytosis is usually a sporadic disorder, but familial forms have been described, often in association with inherited, weakly activating *KIT* mutations.[12,13] Very occasionally, familial clustering of *KIT* D816V has been observed, but in all affected individuals, this mutation is somatically acquired[14] and, as a strongly activating variant, *KIT* D816V is believed to be incompatible with normal embryonic development and thus not transmissible through the germline. Other lines of evidence suggest the possibility of a broader role for genetic variation in mastocytosis. The presence of germline variants in genes known to be somatically mutated in myeloid disorders was one of several factors related to adverse clinical outcome in SM.[11] Studies of mast cell activation disease (MCAD), a disorder that overlaps with SM, indicate a substantial excess of symptoms in first-degree relatives of affected individuals, which might suggest a common genetic susceptibility.[15,16] Several constitutional genetic variants have been associated with the development of different mastocytosis phenotypes in relatively small candidate gene studies[17–21] and a recent single-stage genome-wide association study (GWAS) of 234 affected individuals.[22] Finally, it has been clearly established that constitutional genetic variation at several loci predispose to other myeloproliferative neoplasms (MPN).[23,24]

To determine whether common genetic variation plays a role in predisposition to mastocytosis, we have performed a robust two-stage GWAS focusing on affected individuals that tested positive for *KIT* D816V regardless of clinical subtype to help ensure a genetically homogeneous cohort. We anticipate that the identification of validated genetic markers associated with mastocytosis will provide novel lines of investigation to understand this complex disorder.

Material and methods

Discovery and replication cohorts

Prior to quality control (QC), the stage 1 discovery individuals consisted of 479 *KIT* D816V positive mastocytosis-affected individuals recruited from the UK (n = 329) and Germany (n = 150). These affected individuals were compared with healthy control individuals from the UK Wellcome Trust Case Control Consortium (WTCCC2, n = 5,200)[25] and the German Cooperative Health Research in the Region of Augsburg study (KORA, n = 4,397), respectively.[26] At stage 2, 666 independent *KIT* D816V positive replication individuals were recruited from Spain (n = 399), Denmark (n = 185), and Italy (n = 82) and compared to published population controls from the Spanish National DNA Bank (SNDNAB, n = 1,062),[27,28] a Danish study of ischemic heart disease (Inter99, n = 6,184),[29,30] and the Italian Invecchiare in Chianti study (InCHIANTI, n = 1,210).[31,32] Participants provided informed consent for sampling according to the Declaration of Helsinki. The number of samples that were recruited and used for analysis after QC in the discovery and replication stages is shown in Table S1. An overview of the two-stage study design and sample numbers is shown in Figure S1. All mastocytosis-affected individuals were adults diagnosed via standard procedures. Further details on the five cohorts are provided in the Supplemental methods.[2,4]

Genotyping

DNA was extracted from peripheral blood or bone marrow. The stage 1 affected individuals were genotyped for 960,919 SNPs via Infinium OmniExpress exome chips (version 8_1.4_A1) and the Genome Studio software (GSGT version 1.9.4) at the Clinical Research Facility in Edinburgh. These data are available on request from ArrayExpress (accession number E-MTAB-9358). The stage-2 affected individuals were genotyped for 92 SNPs via custom designed Kompetitive Allele Specific PCR (KASP) at LGC.[33] Genotypic data for the control cohorts were obtained from published studies. In WTCCC2, genotypes were called with Illumina 1.2M Duo chips and Illumina's program to call SNPs with a posterior probability >0.95.[34] KORA control individualss were genotyped for 2,443,177 SNPs via the Illumina human Omni chip (version 2.5-4v1_B) in KORA_A (a subset of follow-up F3 of the population-based survey KORA S3) and 730,372 SNPs with Illumina human Omni express chips (version 12v1_H) in KORA_B (an independent subset of KORA S3/F3). Control individuals from SNDNAB, Inter99, and InCHIANTI were genotyped with Illumina Global Screening arrays, Illumina HumanOmniExpress-24 (versions 1.0A and 1.1A), and Illumina Infinium HumanHap 550K SNP arrays, which include 18, 90, and 45 of the SNPs selected for replication, respectively. Genotypes for the remaining SNPs were determined by imputation.

Quality control

Standard GWAS QC measures[35] were applied to the genotypic data with Plink prior to analysis.[36] These measures included genotype missingness (per sample and per SNP), minor allele frequency (MAF), Hardy Weinberg equilibrium (HWE), heterozygosity (Figure S2), sex inference, cryptic relatedness, strand orientation, and population stratification with multidimensional scaling (MDS) (Figure S3). Since the affected individuals and control individuals were genotyped separately, SNPs were excluded if they had modest deviation from HWE in control individuals (p value $<$ 0.001) or extreme deviation in affected individuals (p value $\leq 1 \times 10^{-10}$), which most likely reflects poor genotyping rather than disease association.[37] The number of SNPs and samples removed by these QC measures is shown in Table S1. QC and imputation of the stage 2 control individuals has previously been described.[28–32] Full details regarding the QC and imputation procedures are given in the Supplemental methods.

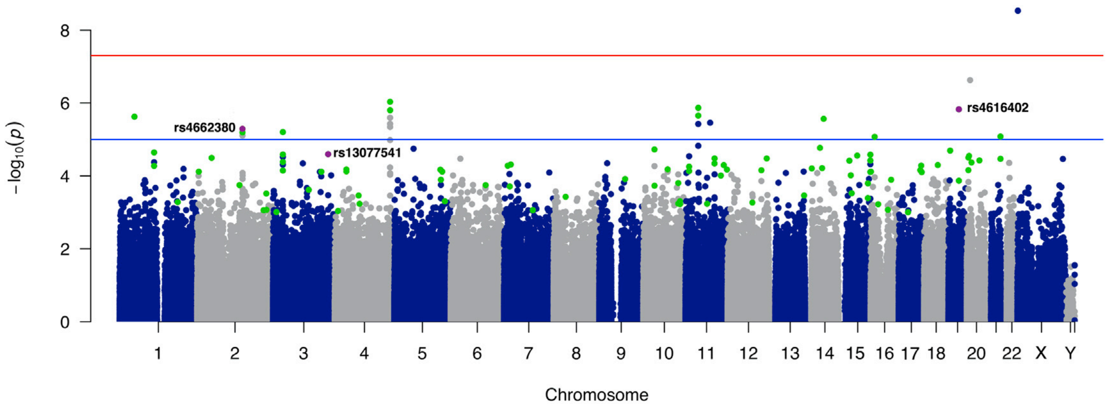

Figure 1. Genome-wide association of *KIT* D816V positive mastocytosis
Manhattan plot showing results from the stage 1 meta-analysis of the UK and German cohorts for all 24 chromosomes. Results are plotted for 592,007 SNPs tested as −log10 of the meta-analysis p values on the y axis against genomic location on the x axis. One SNP was identified with genome-wide significance (p value $< 5 \times 10^{-8}$), indicated by the red line, and a further 18 SNPs were identified with suggestive p values ($<1 \times 10^{-5}$), indicated by the blue line. SNPs selected for replication are highlighted in green, and the three SNPs that reached genome-wide significance after meta-analysis of stages 1 and 2 are highlighted in purple.

Imputation

Imputation of the discovery cohorts was used to increase SNP density and enable fine mapping around significant loci. SNPs were imputed with the Sanger imputation server,[38] which used EAGLE2 for pre-phasing into the Haplotype Reference Consortium (HRC release 1.1) and positional Burrows-Wheeler transform (PBWT) for imputation. Imputed genotypes were quality controlled by exclusion of SNPs with info score <0.80, posterior genotype probabilities less than 0.99, MAF less than 1%, greater than 10% missing genotypes, or extreme deviation from HWE (p value $\leq 1 \times 10^{-10}$).

Statistical analysis

SNPs were tested for association via binary logistic regression in Plink. We carried out a fixed effects inverse variance-weighted meta-analysis by using Plink to combine evidence from the stage 1 cohorts (UK and Germany) and to determine the final effect sizes and significance levels by combining evidence across stages 1 and 2. Heterogeneity between studies was estimated with the χ^2-based Cochran's Q statistic and the I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than chance. To examine the effectiveness of the QC measures and assess evidence for any systematic biases, we used the qqnorm and qqplot procedures in R to construct quantile-quantile (QQ) plots for the stage 1 analysis of the UK and German cohorts and the stage 1 meta-analysis (Figure S4). Samples with evidence of non-Caucasian ancestry were excluded rather than adjusting the association analysis for population stratification. To examine the effect of this decision, we retained the ancestry outliers and repeated the stage 1 analyses with adjustment for the first two principal components from the MDS analysis (Figure S5 and Table S2).

We visualized and interpreted the results from the stage 1 meta-analysis by using the qqman package[39] in R to create a Manhattan plot (Figure 1) and the FUMA software to generate regional plots.[40] Results from the final meta-analysis of stages 1 and 2 were displayed in a forest plot with Stata (Figure 2).

The power to detect SNPs associated with SM was estimated with the genetic power calculator[41] under a multiplicative genetic risk model and a type 1 error rate of 5×10^{-8} (Figure S6). We used a range of genotype relative risks (1.1–2.0) and risk allele frequencies (MAF 0.05–0.4) to estimate power assuming a disease prevalence of 1 in 100,000[42] and unselected control individuals.

Selection of SNPs for replication

To minimize false positives and the potential for overlooking signals with compelling functional evidence but modest significance, we used the following method to select SNPs for follow-up at stage 2. First, we used a clumping procedure in Plink to generate a shortlist of index SNPs ($p < 0.001$) with support from correlated SNPs (SNPs r2 > 0.5, within 500 kb and $p < 0.01$) based on the stage 1 meta-analysis. From this shortlist, 92 index SNPs were selected for replication, and priority, but not exclusivity, was given to SNPs that were either located in or flanked by a gene with functional relevance according to annotation from GeneAlacart.[43] Relevant functions were signal transduction components, hematopoiesis, myeloid leukemia, and myeloproliferative or mast cell conditions from GeneAlacart.[43] A total of 44 SNPs were selected with functional relevance. We then infilled the number of selected SNPs to 82 by selecting the most significant remaining index SNPs. We selected an additional 10 SNPs were selected as backups and to add support to the most promising signals in terms of either their biological relevance, individual significance, or level of support from correlated SNPs.

Identification of chromosomal abnormalities

We identified regions of acquired uniparental disomy (aUPD) and copy number gains or losses in the stage 1 SM-affected individuals by using B allele frequency (BAF) segmentation[44] followed by post processing to select likely somatic events as described[45] and manual review of all BAF plots (Figure S7). See Supplemental methods for further details.

Functional annotation of variants

We explored the biological relevance of regions containing genome-wide significant SNPs by using HaploReg (version 4.1)[46] to annotate the lead SNP and its proxies ($r^2 \geq 0.8$) with respect to histone modification, sequence conservation by using genomic evolutionary rate profiling (GERP),[47] estimated pathogenicity by using combined annotation-dependent depletion (CADD)

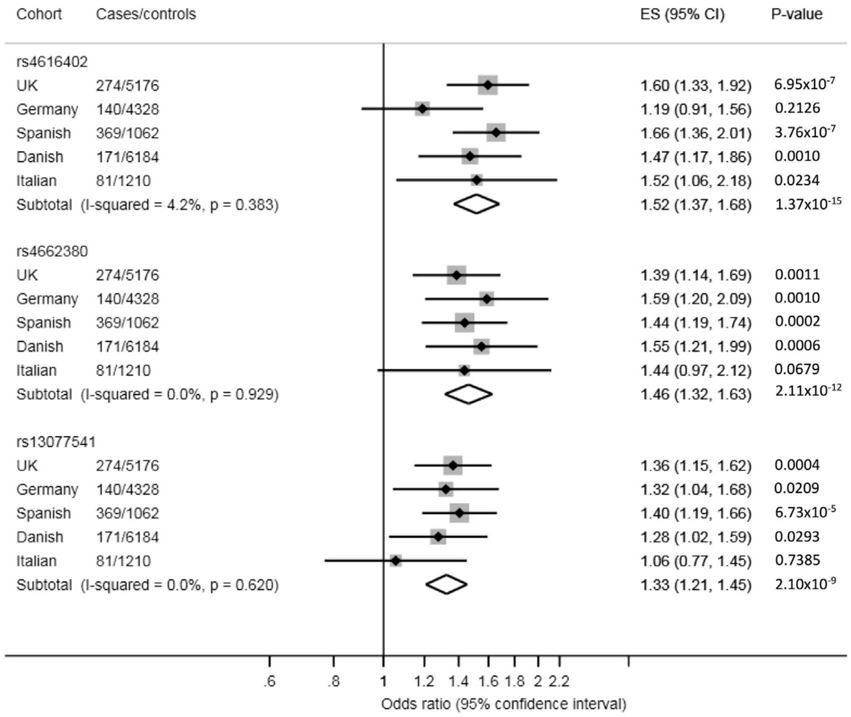

Figure 2. Forest plots and meta-analysis for three SNPs reaching genome-wide significance

Forest plots for each SNP associated with SM at a genome-wide level of significance. Odds ratios (OR = ES) and 95% confidence intervals (CIs) are displayed on the x axis. Results are shown for each cohort (UK, German, Spanish, Danish, and Italian) and the combined analysis. The SNP subtotals and diamond show the final OR and CI for a fixed effects meta-analysis of all five cohorts and uses I^2 to assess heterogeneity in effect sizes between cohorts.

scores,[48] predicted effect on protein binding by using RegulomeDB[49] scores (SNPs scoring ≤3 are likely to affect binding), and previous associations with clinical phenotypes by using the NHGRI-EBI GWAS catalog.[50] Additionally, candidate regions were annotated against a 15-state chromatin model[51] in primary hematopoietic stem cells (E035) and a myeloid leukemia cell line (K562). This model categorizes non-coding DNA into active or repressed states that are respectively enriched and depleted for phenotype-associated SNPs.[52] To gain further functional insight, we performed expression and methylation quantitative trait loci (eQTL and mQTL, respectively) analyses on the lead SNP and its proxies ($r^2 \geq 0.8$) by using GTEx v8[53] and QTLbase.[54] Finally, we used LNCipedia[55] and the Cancer LncRNA Census (CLC)[56] to investigate the function of long non-coding RNA (lncRNA).

Association with clinical features

Diagnostic and phenotypic variables for initial diagnosis (advanced, ASM, SM-AHN, MCL; non-advanced, all other subtypes), the presence or absence of skin lesions (yes or no), gender, baseline serum tryptase (ng/mL), and age were available for most of the Spanish (n = 369) and Italian (n = 81) individuals but not for other cohorts. Three categorical variables (initial diagnosis, skin lesions, and sex) were tested for association with allelic counts for the three significant SNPs via Fisher's exact test. Continuous variables (tryptase and age) were tested via linear regression following Kolmogorov-Smirnov checks for normal distribution and normalization of tryptase levels via quantile transformation. We used a fixed effects inverse variance-weighted meta-analysis to combine evidence from the two cohorts.

Results

Discovery stage

After QC of the stage 1 data, 592,007 SNPs were tested for association with *KIT* D816V positive mastocytosis via binary logistic regression in the UK (274 affected individuals versus 5,176 control individuals) and German cohorts (140 affected individuals versus 4,328 control individuals) (Table S1). Summary statistics from these analyses, which are available from LocusZoom, were combined with a fixed effects meta-analysis.[57] The QQ plots for each analysis and their low genomic inflation factors ($\lambda \leq 1.038$) demonstrate a close agreement with the null hypothesis until the tail of the distribution where SNPs with p values less than 10^{-4} become more significant than expected by chance alone (Figure S4). Consequently, systematic biases such as the separate genotyping of our affected individuals and control individuals, residual population stratification, or clonal somatic changes are unlikely to account for the significance of these SNPs. A Manhattan plot summarizing the results of the stage 1 meta-analysis is shown in Figure 1. A total of 18 SNPs were identified with suggestive p values ($p \leq 1 \times 10^{-5}$).

Replication and final meta-analysis

According to the number of samples that passed QC and using a multiplicative disease model, we estimated the stage 1 analysis to have 80% power to detect common SNPs (MAF = 0.4) with a relative risk (RR) of 1.56 and rare SNPs (MAF = 0.1) with an RR of 1.82 (Figure S6A). Because of the potential to overlook SNPs with smaller effect sizes, we used a set of selection criteria rather than significance alone (see Material and methods) to identify 92 SNPs for replication. These SNPs were selected to have support from correlated SNPs and were either the most significant (n = 38), surpassed a moderate significance threshold ($p < 0.001$) and were located in or flanked by a functionally relevant gene (n = 44), or were selected as backups for the most promising signals (n = 10). One SNP, rs7884433, achieved genome-wide significance in the stage 1 analysis, but it was not selected for replication because it lacked support from any of the SNPs in strong linkage disequilibrium (LD) and is thus likely to be a technical artifact.

Table 1. Summary of the most significant SNPs from meta-analysis of stages 1 and 2

SNP	Chr	Location (hg19)	Alleles	RAF	Gene	p_{meta}	OR (CI)	I^2
rs4616402	19q13	33,753,555	A/G	0.240	*SLC7A10-CEBPA*	1.37×10^{-15}	1.52 (1.37–1.68)	4.2
rs4662380	2q22	145,316,407	C/T	0.189	*LINC01412*	2.11×10^{-12}	1.46 (1.32–1.63)	0
rs13077541	3q26	176,925,740	G/A	0.464	*TBL1XR1-LINC00501*	2.10×10^{-9}	1.33 (1.21–1.45)	0

SNP, rs identifier from dbSNP; alleles, risk associated/non-risk associated allele; RAF, risk allele frequency in Europeans from 1000 genomes; p_{meta}, fixed effects meta-analysis of stages 1 and 2; OR, odds ratio; CI, 95% confidence interval; I^2, heterogeneity index (0–100).

Of the 92 SNPs selected, 75 were successfully genotyped in 666 *KIT* D816V mastocytosis-affected individuals from Spain, Denmark, and Italy. Additional control individuals (n = 8,456) from the same populations that had previously been genotyped were used for comparison. After QC, 621 affected individuals and all the control individuals remained for analysis. All SNPs passed QC in affected individuals, although 19 were excluded from the Spanish control individuals because of per SNP missingness ($\geq$10%) following imputation. Samples were tested for association with SM as three separate cohorts via binary logistic regression. We determined the final significance levels and effect sizes by using a fixed effects inverse variance-weighted meta-analysis to combine evidence from stages 1 and 2. This meta-analysis identified three intergenic SNPs with genome-wide significance: rs4616402 ($p_{meta} = 1.37 \times 10^{-15}$), rs4662380 ($p_{meta} = 2.11 \times 10^{-12}$), and rs13077541 ($p_{meta} = 2.10 \times 10^{-9}$) (Table 1). Results for the three SNPs reaching genome-wide significance are summarized in a forest plot that shows that each SNP is significant in four of the five cohorts tested and that there is evidence for the same trend in the remaining population (Figure 2). Cochran's Q test and I^2 statistics showed that for each SNP there was no evidence of heterogeneity between cohorts. Results from the meta-analysis of stages 1 and 2 for all SNPs tested are shown in Table S3.

To investigate the possibility of residual population stratification, we repeated the stage 1 analyses without removing 26 samples with evidence of outlying ancestry (Table S1) and adjusting the association analysis by using the first two principal components from MDS. The top three SNPs retained genome-wide significance, and rs4662380 and rs13077541 became slightly more significant (Table S2), which suggests an absence of residual population stratification in the original analysis.

Functional annotation and candidate gene mapping

To explore the functional relevance of the regions associated with mastocytosis, we used HaploReg and RegulomeDB to determine whether the risk SNP or its proxies ($r^2 \geq 0.8$) were located in regions with potential regulatory functions based on chromatin modification, DNA methylation, and alteration of transcription factor (TF)-binding motifs (Table S4). To gain further functional insight, we performed eQTL and mQTL analyses on the lead SNP and its proxies by using GTEx v8[53] and QTLbase.[54] Finally, we repeated the stage 1 meta-analysis by using imputation to enable fine mapping around the lead SNPs and to generate association results for proxies, which had not been directly genotyped.

The most significant SNP, rs4616402, confers a 1.52-fold increased risk of developing mastocytosis and is situated in an intergenic region on chromosome 19 between a solute carrier gene (*SLC7A10*, 36.8 kb downstream) and a gene encoding a transcription factor (*CEBPA*, 37.2 kb downstream) that coordinates proliferation and differentiation of myeloid progenitor cells (Figure 3A). Using QTLbase, we found that rs4616402 is strongly associated with the expression of *CEBPA* in whole blood according to data from three previous eQTL studies ($p_{eQTL} = 2.30 \times 10^{-14}$; $p_{eQTL} = 2.96 \times 10^{-11}$; $p_{eQTL} = 9.20 \times 10^{-9}$).[58–60] There is no evidence that *SLC7A10* has a role in carcinogenesis, including myeloid malignancies, and no additional SNPs were identified in strong LD with rs4616402. However, there is weak evidence that rs4616402 may have functional consequences according to the RegulomeDB score (score = 4). The chromatin surrounding rs4616402 is characterized as an enhancer (7_Enh) in primary hematopoietic stem cells because of an enrichment of the H3K4me1 signature. Additionally, the risk allele is predicted to alter three TF-binding motifs (Arnt_1, Gm397, and Hmx_1, Table S4).

The second most significant SNP, rs4662380, increases the risk of developing mastocytosis by 1.46-fold and is located in the first intron of a lincRNA gene (*LINC01412*) (Figure 3B). Twelve additional SNPs in LINC01412 were identified in strong LD with the lead. Three of these proxies are located in chromatin enhancers (7_Enh: rs6722387, rs16823865, and rs13413446) in primary hematopoietic stem cells, and one is located in a flanking active transcription start site (2_TssAFlnk: rs16823855) in K562 (Table S4). The RegulomeDB scores indicate that two of the proxies, rs4662227 (score = 2c) and rs13413446 (score = 3a), are likely to affect TF binding, while the remaining SNPs are estimated to have weak evidence for functional consequences. However, using the GWAS catalog,[50] we found that one of the remaining proxies, rs16823866, was strongly associated with white blood cell counts in two previous studies ($p = 4 \times 10^{-18}$ and $p = 6 \times 10^{-11}$).[62,63] Finally, using QTLbase, we found that the lead SNP ($p_{eQTL} = 2.55 \times 10^{-11}$) and four proxies, including rs16823866 ($p_{eQTL} = 2.55 \times 10^{-11}$), were strongly associated with the expression of the nearby gene *TEX41* in neutrophils.[64]

The final SNP, rs13077541, is associated with a 1.33-fold increase in risk of developing mastocytosis and is located

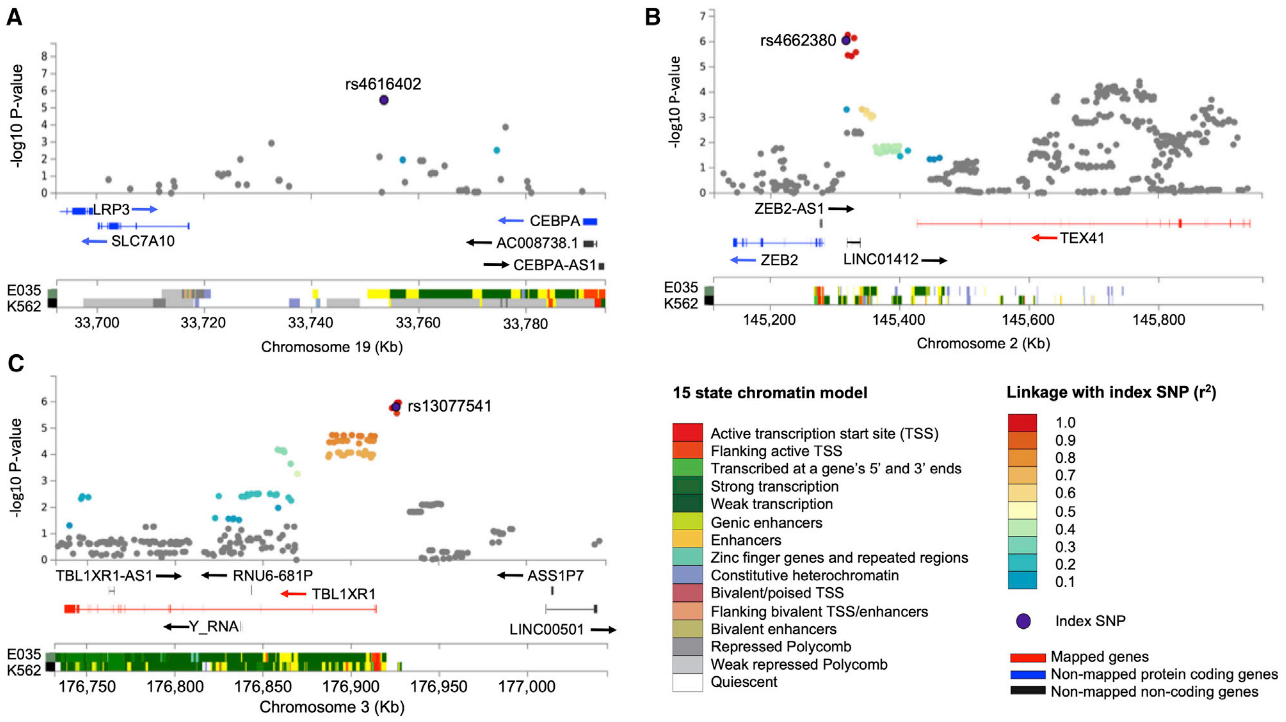

Figure 3. Regional plots of the imputed stage 1 meta-analysis for SNPs reaching genome-wide significance in the final meta-analysis
(A–C) Results from the imputed stage 1 meta-analysis in a region surrounding three SNPs (rs4616402 [A], rs4662380 [B], and rs13077541 [C]) that predispose to SM and reached genome-wide significance in the final meta-analysis. In each plot, the leading SNP is indicated by a purple circle and the color of other SNPs represent the strength of linkage disequilibrium (r^2) with the lead SNP. Protein-coding genes and RNA genes are shown in the track below with arrows to indicate the direction of transcription and wider lines representing the location of exons. The lower panel displays the 15-state chromatin track (chromHMM) in primary hematopoietic stem cells (E035) and K562 with data from the NIH Roadmap Epigenomics Consortium.[61] Physical positions are relative to build 37 (hg19) of the human genome.

in an intergenic region of chromosome 3 between transducin (β)-like 1 X-linked receptor 1 (*TBL1XR1*, 10.6 kb upstream) and another lincRNA gene (*LINC00501*, 86.5 kb upstream) (Figure 3C). Fifty-three additional SNPs were identified in strong LD with the lead, including 27 intronic SNPs in *TBL1XR1* (Table S4). Eleven of these proxies are located in active chromatin regions, including three in an active transcription start site (1_TssA: rs12493005, rs12486557, and rs34302523) and two in a 5′ transcribed region (3_TxFlnk: rs35072945 and rs34311793) in K562. The RegulomeDB scores indicate that five of the proxies are likely to affect binding (score2a–c: rs6790639, rs34302523, rs6772872, rs7616138, and rs1920131). Of these, rs6790639 is particularly relevant because the PU.1 TF, which is encoded by the Spi-1 proto-oncogene (*SPI1*), has been shown to bind to this region in K562 via ChIP sequencing.[65] PU.1, together with other TFs, regulates the expression of genes involved myelopoiesis.[66] Using QTLbase, we found that the lead SNP ($p_{eQTL} = 5.70 \times 10^{-8}$) and one of the proxies, rs16823866 ($p_{eQTL} = 9.52 \times 10^{-9}$), were strongly associated with the expression of *TBL1XR1* in CD4+ naive T cells.[64]

Association with clinical features

To determine whether variants that predispose to the development of mastocytosis relate to particular clinical features, we used Fisher's exact tests and linear regression to correlate allelic counts for the three significant SNPs with clinical phenotypes in the Spanish and Italian cohorts (Table 2), the only affected individuals for which clinical information was available. A significant association that remained significant after correction for multiple testing was identified between rs4616402 and age at presentation (n = 422; p = 0.009; beta = 4.41) in individuals with non-advanced disease. No association with age was seen in the much smaller group of individuals (n = 26) with advanced disease, a subgroup for which additional mutations may be a confounding factor. In affected individuals, the age of onset was estimated to increase by 4.41 years per risk allele. No associations were seen with baseline tryptase levels, gender, skin lesions, or disease phenotype.

Association with *TPSAB1* and *TPSB2*

Increased copy number variation at *TPSAB1*, the gene at 16p13 encoding α-tryptase, is associated with elevated serum tryptase levels in hereditary α-tryptasemia.[67] Our analysis did not include direct copy number analysis of this gene; however, a recent study linked *TPSAB1* duplications with three SNPs, including rs58124832.[68] This SNP was genotyped at stage 1 and met our criteria for analysis at stage 2, yielding a suggestive overall association with SM ($p_{meta} = 9.03 \times 10^{-6}$). The Cochran's Q test and I^2 statistics showed no evidence of heterogeneity between

Table 2. Association between the most significant SNPs and clinical phenotypes in the Spanish and Italian cohorts

Phenotype	Number of affected individuals	rs4662380		rs13077541		rs4616402	
		p value	Effect size (CI)	p value	Effect size (CI)	p value	Effect size (CI)
Initial diagnosis (indolent/advanced)	422/26	0.175	0.58 (0.26–1.27)	0.646	0.88 (0.50–1.54)	0.238	0.60 (0.25–1.40)
Sex (F/M)	235/214	0.266	1.18 (0.88–1.60)	0.384	1.12 (0.86–1.46)	0.904	1.03 (0.65–1.61)
Skin lesions (+/−)	275/122	0.638	1.08 (0.77–1.51)	0.151	0.81 (0.60–1.08)	0.406	1.23 (0.75–2.00)
Age at diagnosis	422	0.668	0.55 (−1.97–3.07)	0.625	0.67 (−2.02–3.35)	0.009	4.41 (1.09–7.73)
Tryptase	417	0.452	−0.08 (−0.29–0.13)	0.136	−0.17 (−0.39–0.05)	0.249	0.17 (−0.12–0.45)

Categorical phenotypes: initial diagnosis (422 indolent versus 26 advanced mastocytosis-affected individuals), sex (235 female versus 214 male individuals), and skin lesions (275 individuals with skin phenotype versus 122 individuals without skin phenotype); p value, fixed effects meta-analysis of Italian and Spanish Fisher's exact test; effect size, odds ratio; CI, 95% confidence interval. Continuous phenotypes: age at diagnosis and tryptase levels tested in individuals with non-advanced phenotype; p value, linear regression; effect size, regression coefficient beta; CI, 95% confidence interval.

cohorts; however, the association was significant in only three cohorts ($p_{German} = 0.0058$, $p_{UK} = 0.0042$, and $p_{Spanish} = 0.05$). The eQTL analysis showed that rs58124832 is strongly associated with the expression of *TPSAB1* ($p_{eQTL} < 1.9 \times 10^{-58}$) and *TPSB2* (tryptase-β2; $p_{eQTL} = 1.96 \times 10^{-75}$) in blood.

Association with *TERT*

Several *TERT* SNPs have been identified as risk factors for the development of hematological malignancies, including MPN, as well as some solid tumors. Our stage 1 analysis included rs2853677, which has been linked to both MPN and *JAK2* V617F associated clonal hematopoiesis.[24] This SNP marginally failed to meet our criteria for analysis at stage 2; however, the stage 1 meta-analysis for directly genotyped UK and German affected individuals showed $p_{meta} = 0.0011$, suggesting the possibility of an association. To examine this in more detail, we imputed genotypes for 64 additional SNPs spanning *TERT* and tested their association with SM. As shown in Table S5, seven SNPs achieved p values < 0.001. The strongest of these was for rs7726159 ($p_{meta} = 8 \times 10^{-5}$), an established risk SNP for multiple cancer types.[69] We identified one secondary association at *TERT* for rs2853677, which remained significant after conditioning on rs7726159 ($p_{conditional} = 0.035$). No associations were seen with other SNPs that predispose to MPN[70] or clonal hematopoiesis of indeterminate potential[71] in our stage 1 data (Table S6).

Associations with other genetic factors

To the best of our knowledge, 14 SNPs have been associated with the development or phenotype of human mastocytosis in published studies.[17–22] Of these, 11 were directly genotyped or could be imputed from our stage 1 data (Table S7), but only one of these was significant: rs1800925 in the promoter region of *IL13* at 5q31 ($p_{imputed} = 0.008$). This SNP has been linked to the development of adult SM and serum interleukin-13 levels[18] and inflammatory disorders such as chronic obstructive pulmonary disease.[72]

Discussion

Despite being characterized by a common somatic oncogenic driver mutation, mastocytosis is a complex disorder with a broad range of clinical phenotypes and outcomes. In this study, we have identified constitutional genotype as an additional factor contributing to the heterogeneity of mastocytosis. The use of a molecular definition for affected individuals rather than clinically defined subtypes and careful ethnicity matching of affected individuals and control individuals aimed to reduce the chance of heterogeneity both in the primary and replication cohorts. Thus, with a relatively modest cohort size for a GWAS, we were able to identify and validate three novel SNPs that achieved genome-wide significance and additional suggestive associations at *TERT*, *TPSAB1/TPSB2*, and *IL13* that merit further investigation. Notably, apart from rs1800925 (*IL13*), we did not confirm any of the previously published associations derived from candidate gene studies and a recent GWAS that did not include a replication cohort (Table S6). In addition, we found no evidence that genetic variation at *KIT* is associated with acquisition of *KIT* D816V, unlike the finding in MPN that the *JAK2* haplotype strongly influences the probability of acquiring *JAK2* V617F.[73]

Theoretically, common genetic variation may influence mastocytosis by distinct mechanisms, for example by promoting or favoring the outgrowth of a *KIT* D816V positive clone that arose by random mutation (fertile ground hypothesis); by increasing the probability that a *KIT* D816V mutation arises in a stem cell (hypermutability hypothesis); or by promoting the development of signs or symptoms in an individual with a *KIT* D816V positive clone, thus increasing the chance of clinical investigation (phenotypic hypothesis). We considered the possibility that clonal somatic changes might affect the analysis; however, we found that mastocytosis genomes are relatively simple in that only a small proportion of affected individuals showed likely somatic copy number changes or acquired uniparental disomy (Figure S7). Furthermore, apart from isolated affected individuals, the genomic regions with somatic changes did not include the risk factors we identified.

Of the three significant SNPs identified in this study, the strongest association was seen for rs4616402 at 19q13. Interestingly, this SNP was significantly associated with age of diagnosis in individuals with non-advanced disease. This SNP is located in a candidate enhancer, and the risk allele is linked to reduced expression of *CEBPA*,[60] located 37.3 kb upstream. Another 19q13 SNP, rs78744187, has previously been linked to basophil counts and shown to modulate the activity of a *CEBPA* enhancer;[74] however, this variant is not in LD with rs4616402 ($r^2 = 0.22$). *CEBPA* is an intronless gene that encodes a leucine zipper TF that binds to the CCAAT motif in the promoter of its target genes. It is expressed in myeloid progenitor cells, and several studies have defined its critical role in myelopoiesis and malignant transformation of myeloid cells.[75] Of particular relevance, high C/EBPα expression inhibits the production of mast cells from mast/basophil common progenitors, whereas low C/EBPα expression inhibits the production of basophils.[71] Although the consequence of reduced *CEBPA* levels in the context of *KIT* D816V remains to be defined, reduced *CEBPA* expression associated with rs4616402 may be relevant to the fertile ground and phenotypic hypothesis defined above by creating an environment that favors the production of mast cells. It is striking that *CEBPA* or its product, C/EBPα, is targeted by two other oncogenic tyrosine kinases: BCR-ABL1 downregulates *CEBPA* by a post-transcriptional mechanism[76] and oncogenic FLT3 mutants disrupt C/EBPα function by ERK1/2-mediated phosphorylation.[77] Furthermore, low *CEBPA* expression is commonly seen in acute myeloid leukemia, although the underlying mechanism is unclear.[75] Detailed functional studies are needed to clarify the relationship between *KIT* D816V-driven clonal outgrowth and *CEBPA* expression.

The second most significant SNP, rs4662380, is located at 2q22 within the lincRNA *LINC01412* and associated with higher expression of the nearby gene *TEX41*. Both are of unknown function, but because of the possibility of long range interactions between GWAS signals and target genes, it is unclear whether either are directly relevant to SM. *ZEB2* is another nearby gene that has been linked to both myeloid and lymphoid leukemias,[78,79] but we found no association between rs4662380 and *ZEB2* expression. Interestingly, rs16823866, a SNP strongly linked to rs4662380, was associated with elevated white blood cells and, specifically, basophils in three independent population studies.[62,63,80] Although the underlying mechanism is unclear, this may be relevant to the phenotypic hypothesis in that affected individuals with abnormal blood counts may be more likely to be investigated clinically. The final SNP, rs13077541, is linked to expression of *TBL1XR1*. This gene has been reported as a fusion partner of *PDGFRB*, *ROS1*, *RARA*, and *RARB* in myeloid malignancies,[81–83] but its significance in relation to SM remains to be established.

Data and code availability

Genotyping data are available at ArrayExpress (https://www.ebi.ac.uk/arrayexpress/; accession number E-MTAB-9358). GWAS summary statistics are available at LocusZoom (http://locuszoom.org/ under "Mastocytosis GWAS").

Supplemental Information

Supplemental Information can be found online at https://doi.org/10.1016/j.ajhg.2020.12.007.

Acknowledgments

A full list of the investigators who contributed to the generation of the WTCCC data is available from the WTCCC website, funding for which was provided by The Wellcome Trust under award 07611. The KORA study was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The KORA Study Group consists of A. Peters (speaker), J. Heinrich, R. Holle, R. Leidl, C. Meisinger, K. Strauch, and their co-workers, who are responsible for the design and conduct of the KORA studies. We gratefully acknowledge the contribution of all members of field staff conducting the KORA study, and we are grateful to all study participants of KORA for their invaluable contributions to this study. The InCHIANTI study baseline (1998–2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (263 MD 9164 and 263 MD 821336). The Spanish mastocytosis cohort and the Spanish National DNA Bank were supported by grants from the Instituto de Salud Carlos III and FEDER (PI16/00642 and PT17/0015/0044). The Italian cohort of mastocytosis-affected individuals was funded by the Associazione Italiana per la Ricerca sul cancro, Mynerva project, 21267. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation. A.A.Z.D. was supported by a Lady Tata International Award; G.G., N.C.P.C., Y.H., A.J.C., and W.J.T. were supported by Blood Cancer UK (13002 and 18007).

Declaration of interests

The authors declare no competing interests

Received: October 9, 2020
Accepted: December 7, 2020
Published: January 8, 2021

Web resources

OMIM, https://www.omim.org/entry/154800

Wellcome Trust Case Control Consortium, https://www.wtccc.org.uk/

References

1. Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391–2405.
2. Valent, P., Akin, C., and Metcalfe, D.D. (2017). Mastocytosis: 2016 updated WHO classification and novel emerging treatment concepts. Blood *129*, 1420–1427.
3. Cohen, S.S., Skovbo, S., Vestergaard, H., Kristensen, T., Møller, M., Bindslev-Jensen, C., Fryzek, J.P., and Broesby-Olsen, S. (2014). Epidemiology of systemic mastocytosis in Denmark. Br. J. Haematol. *166*, 521–528.
4. Arock, M., Sotlar, K., Akin, C., Broesby-Olsen, S., Hoermann, G., Escribano, L., Kristensen, T.K., Kluin-Nelemans, H.C., Hermine, O., Dubreuil, P., et al. (2015). KIT mutation analysis in mast cell neoplasms: recommendations of the European Competence Network on Mastocytosis. Leukemia *29*, 1223–1232.
5. Sperr, W.R., Kundi, M., Alvarez-Twose, I., van Anrooij, B., Oude Elberink, J.N.G., Gorska, A., Niedoszytko, M., Gleixner, K.V., Hadzijusufovic, E., Zanotti, R., et al. (2019). International prognostic scoring system for mastocytosis (IPSM): a retrospective cohort study. Lancet Haematol. *6*, e638–e649.
6. Erben, P., Schwaab, J., Metzgeroth, G., Horny, H.P., Jawhar, M., Sotlar, K., Fabarius, A., Teichmann, M., Schneider, S., Ernst, T., et al. (2014). The KIT D816V expressed allele burden for diagnosis and disease monitoring of systemic mastocytosis. Ann. Hematol. *93*, 81–88.
7. Hoermann, G., Gleixner, K.V., Dinu, G.E., Kundi, M., Greiner, G., Wimazal, F., Hadzijusufovic, E., Mitterbauer, G., Mannhalter, C., Valent, P., and Sperr, W.R. (2014). The KIT D816V allele burden predicts survival in patients with mastocytosis and correlates with the WHO type of the disease. Allergy *69*, 810–813.
8. Muñoz-González, J.I., Álvarez-Twose, I., Jara-Acevedo, M., Henriques, A., Viñas, E., Prieto, C., Sánchez-Muñoz, L., Caldas, C., Mayado, A., Matito, A., et al. (2019). Frequency and prognostic impact of *KIT* and other genetic variants in indolent systemic mastocytosis. Blood *134*, 456–468.
9. Jawhar, M., Schwaab, J., Álvarez-Twose, I., Shoumariyeh, K., Naumann, N., Lübke, J., Perkins, C., Muñoz-González, J.I., Meggendorfer, M., Kennedy, V., et al. (2019). MARS: Mutation-Adjusted Risk Score for Advanced Systemic Mastocytosis. J. Clin. Oncol. *37*, 2846–2856.
10. Jawhar, M., Schwaab, J., Schnittger, S., Meggendorfer, M., Pfirrmann, M., Sotlar, K., Horny, H.P., Metzgeroth, G., Kluger, S., Naumann, N., et al. (2016). Additional mutations in SRSF2, ASXL1 and/or RUNX1 identify a high-risk group of patients with KIT D816V(+) advanced systemic mastocytosis. Leukemia *30*, 136–143.
11. Muñoz-González, J.I., Jara-Acevedo, M., Alvarez-Twose, I., Merker, J.D., Teodosio, C., Hou, Y., Henriques, A., Roskin, K.M., Sanchez-Muñoz, L., Tsai, A.G., et al. (2018). Impact of somatic and germline mutations on the outcome of systemic mastocytosis. Blood Adv. *2*, 2814–2828.
12. Zhang, L.Y., Smith, M.L., Schultheis, B., Fitzgibbon, J., Lister, T.A., Melo, J.V., Cross, N.C., and Cavenagh, J.D. (2006). A novel K509I mutation of KIT identified in familial mastocytosis-in vitro and in vivo responsiveness to imatinib therapy. Leuk. Res. *30*, 373–378.
13. Wasag, B., Niedoszytko, M., Piskorz, A., Lange, M., Renke, J., Jassem, E., Biernat, W., Debiec-Rychter, M., and Limon, J. (2011). Novel, activating KIT-N822I mutation in familial cutaneous mastocytosis. Exp. Hematol. *39*, 859–65.e2.
14. Zanotti, R., Simioni, L., Garcia-Montero, A.C., Perbellini, O., Bonadonna, P., Caruso, B., Jara-Acevedo, M., Bonifacio, M., and De Matteis, G. (2013). Somatic D816V KIT mutation in a case of adult-onset familial mastocytosis. J. Allergy Clin. Immunol. *131*, 605–607.
15. Molderings, G.J., Haenisch, B., Bogdanow, M., Fimmers, R., and Nöthen, M.M. (2013). Familial occurrence of systemic mast cell activation disease. PLoS ONE *8*, e76241.
16. Haenisch, B., Nöthen, M.M., and Molderings, G.J. (2012). Systemic mast cell activation disease: the role of molecular genetic alterations in pathogenesis, heritability and diagnostics. Immunology *137*, 197–205.
17. Daley, T., Metcalfe, D.D., and Akin, C. (2001). Association of the Q576R polymorphism in the interleukin-4 receptor alpha chain with indolent mastocytosis limited to the skin. Blood *98*, 880–882.
18. Nedoszytko, B., Niedoszytko, M., Lange, M., van Doormaal, J., Gleń, J., Zabłotna, M., Renke, J., Vales, A., Buljubasic, F., Jassem, E., et al. (2009). Interleukin-13 promoter gene polymorphism -1112C/T is associated with the systemic form of mastocytosis. Allergy *64*, 287–294.
19. Rausz, E., Szilágyi, A., Nedoszytko, B., Lange, M., Niedoszytko, M., Lautner-Csorba, O., Falus, A., Aladzsity, I., Kokai, M., Valent, P., et al. (2013). Comparative analysis of IL6 and IL6 receptor gene polymorphisms in mastocytosis. Br. J. Haematol. *160*, 216–219.
20. Lange, M., Gleń, J., Zabłotna, M., Nedoszytko, B., Sokołowska-Wojdyło, M., Rębała, K., Ługowska-Umer, H., Niedoszytko, M., Górska, A., Sikorska, M., et al. (2017). Interleukin-31 Polymorphisms and Serum IL-31 Level in Patients with Mastocytosis: Correlation with Clinical Presen-tation and Pruritus. Acta Derm. Venereol. *97*, 47–53.
21. Nedoszytko, B., Lange, M., Renke, J., Niedoszytko, M., Zabłotna, M., Gleń, J., and Nowicki, R. (2018). The Possible Role of Gene Variant Coding Nonfunctional Toll-Like Receptor 2 in the Pathogenesis of Mastocytosis. Int. Arch. Allergy Immunol. *177*, 80–86.
22. Nedoszytko, B., Sobalska-Kwapis, M., Strapagiel, D., Lange, M., Górska, A., Elberink, J.N.G.O., van Doormaal, J., Słomka, M., Kalinowski, L., Gruchała-Niedoszytko, M., et al. (2020). Results from a Genome-Wide Association Study (GWAS) in Mastocytosis Reveal New Gene Polymorphisms Associated with WHO Subgroups. Int. J. Mol. Sci. *21*, 5506.
23. Tapper, W., Jones, A.V., Kralovics, R., Harutyunyan, A.S., Zoi, K., Leung, W., Godfrey, A.L., Guglielmelli, P., Callaway, A., Ward, D., et al. (2015). Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat. Commun. *6*, 6691.
24. Hinds, D.A., Barnholt, K.E., Mesa, R.A., Kiefer, A.K., Do, C.B., Eriksson, N., Mountain, J.L., Francke, U., Tung, J.Y., Nguyen, H.M., et al. (2016). Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood *128*, 1121–1128.
25. International Parkinson's Disease Genomics Consortium (IPDGC); and Wellcome Trust Case Control Consortium 2 (WTCCC2) (2011). A two-stage meta-analysis identifies several new loci for Parkinson's disease. PLoS Genet. *7*, e1002142.
26. Wichmann, H.E., Gieger, C., Illig, T.; and MONICA/KORA Study Group (2005). KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl 1*), S26–S30.

27. Bosch, X. (2004). Spain to establish national genetic database. Lancet *363*, 1044.

28. Julià, A., Domènech, E., Ricart, E., Tortosa, R., García-Sánchez, V., Gisbert, J.P., Nos Mateu, P., Gutiérrez, A., Gomollón, F., Mendoza, J.L., et al. (2013). A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. Gut *62*, 1440–1445.

29. Jørgensen, T., Borch-Johnsen, K., Thomsen, T.F., Ibsen, H., Glümer, C., and Pisinger, C. (2003). A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. Eur. J. Cardiovasc. Prev. Rehabil. *10*, 377–386.

30. Pisinger, C., Vestbo, J., Borch-Johnsen, K., and Jørgensen, T. (2005). Smoking cessation intervention in a large randomised population-based study. The Inter99 study. Prev. Med. *40*, 285–292.

31. Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T.B., and Guralnik, J.M. (2000). Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J. Am. Geriatr. Soc. *48*, 1618–1625.

32. Tanaka, T., Shen, J., Abecasis, G.R., Kisialiou, A., Ordovas, J.M., Guralnik, J.M., Singleton, A., Bandinelli, S., Cherubini, A., Arnett, D., et al. (2009). Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. PLoS Genet. *5*, e1000338.

33. He, C., Holme, J., and Anthony, J. (2014). SNP genotyping: the KASP assay. Methods Mol. Biol. *1145*, 75–86.

34. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

35. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat. Protoc. *5*, 1564–1573.

36. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

37. Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E.M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int. J. Methods Psychiatr. Res. *27*, e1608.

38. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

39. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. https://doi.org/10.1101/005165.

40. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.

41. Purcell, S., Cherny, S.S., and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics *19*, 149–150.

42. Coltoff, A., and Mascarenhas, J. (2019). Relevant updates in systemic mastocytosis. Leuk. Res. *81*, 10–18.

43. Stelzer, G., Dalah, I., Stein, T.I., Satanower, Y., Rosen, N., Nativ, N., Oz-Levi, D., Olender, T., Belinky, F., Bahir, I., et al. (2011). In-silico human genomics with GeneCards. Hum. Genomics *5*, 709–717.

44. Staaf, J., Lindgren, D., Vallon-Christersson, J., Isaksson, A., Göransson, H., Juliusson, G., Rosenquist, R., Höglund, M., Borg, A., and Ringnér, M. (2008). Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. Genome Biol. *9*, R136.

45. Dawoud, A.A.Z., Tapper, W.J., and Cross, N.C.P. (2020). Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. Leukemia *34*, 2660–2672.

46. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44* (D1), D877–D881.

47. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., Sidow, A.; and NISC Comparative Sequencing Program (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.

48. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

49. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.

50. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (D1), D1005–D1012.

51. Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

52. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. *41*, 827–841.

53. Carithers, L.J., and Moore, H.M. (2015). The Genotype-Tissue Expression (GTEx) Project. Biopreserv. Biobank. *13*, 307–308.

54. Zheng, Z., Huang, D., Wang, J., Zhao, K., Zhou, Y., Guo, Z., Zhai, S., Xu, H., Cui, H., Yao, H., et al. (2020). QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Res. *48* (D1), D983–D991.

55. Volders, P.J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., and Vandesompele, J. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res. *47* (D1), D135–D139.

56. Carlevaro-Fita, J., Lanzós, A., Feuerbach, L., Hong, C., Mas-Ponte, D., Pedersen, J.S., Johnson, R.; PCAWG Drivers and Functional Interpretation Group; and PCAWG Consortium (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. Commun. Biol. *3*, 56.

57. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

58. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S.,

et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. https://doi.org/10.1101/447367.

59. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.
60. Lloyd-Jones, L.R., Holloway, A., McRae, A., Yang, J., Small, K., Zhao, J., Zeng, B., Bakshi, A., Metspalu, A., Dermitzakis, M., et al. (2017). The Genetic Architecture of Gene Expression in Peripheral Blood. Am. J. Hum. Genet. *100*, 228–237.
61. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
62. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat. Genet. *50*, 390–400.
63. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell *167*, 1415–1429.e19.
64. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398–1414.e24.
65. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
66. van Riel, B., and Rosenbauer, F. (2014). Epigenetic control of hematopoiesis: the PU.1 chromatin connection. Biol. Chem. *395*, 1265–1274.
67. Lyons, J.J., Yu, X., Hughes, J.D., Le, Q.T., Jamil, A., Bai, Y., Ho, N., Zhao, M., Liu, Y., O'Connell, M.P., et al. (2016). Elevated basal serum tryptase identifies a multisystem disorder associated with increased TPSAB1 copy number. Nat. Genet. *48*, 1564–1569.
68. Lyons, J.J., Stotz, S.C., Chovanec, J., Liu, Y., Lewis, K.L., Nelson, C., DiMaggio, T., Jones, N., Stone, K.D., Sung, H., et al. (2018). A common haplotype containing functional CACNA1H variants is frequently coinherited with increased TPSAB1 copy number. Genet. Med. *20*, 503–512.
69. Wang, Z., Zhu, B., Zhang, M., Parikh, H., Jia, J., Chung, C.C., Sampson, J.N., Hoskins, J.W., Hutchinson, A., Burdette, L., et al. (2014). Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. Hum. Mol. Genet. *23*, 6616–6633.
70. Bao, E.L., Nandakumar, S.K., Liao, X., Bick, A.G., Karjalainen, J., Tabaka, M., Gan, O.I., Havulinna, A.S., Kiiskinen, T.T.J., Lareau, C.A., et al.; FinnGen; and 23andMe Research Team (2020). Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. Nature *586*, 769–775.
71. Bick, A.G., Weinstock, J.S., Nandakumar, S.K., Fulco, C.P., Bao, E.L., Zekavat, S.M., Szeto, M.D., Liao, X., Leventhal, M.J., Nasser, J., et al.; NHLBI Trans-Omics for Precision Medicine Consortium (2020). Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature *586*, 763–768.
72. Ahmadi, A., Ghaedi, H., Salimian, J., Azimzadeh Jamalkandi, S., and Ghanei, M. (2019). Association between chronic obstructive pulmonary disease and interleukins gene variants: A systematic review and meta-analysis. Cytokine *117*, 65–71.
73. Jones, A.V., Chase, A., Silver, R.T., Oscier, D., Zoi, K., Wang, Y.L., Cario, H., Pahl, H.L., Collins, A., Reiter, A., et al. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat. Genet. *41*, 446–449.
74. Guo, M.H., Nandakumar, S.K., Ulirsch, J.C., Zekavat, S.M., Buenrostro, J.D., Natarajan, P., Salem, R.M., Chiarle, R., Mitt, M., Kals, M., et al. (2017). Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. Proc. Natl. Acad. Sci. USA *114*, E327–E336.
75. Avellino, R., and Delwel, R. (2017). Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation. Blood *129*, 2083–2091.
76. Perrotti, D., Cesi, V., Trotta, R., Guerzoni, C., Santilli, G., Campbell, K., Iervolino, A., Condorelli, F., Gambacorti-Passerini, C., Caligiuri, M.A., and Calabretta, B. (2002). BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat. Genet. *30*, 48–58.
77. Radomska, H.S., Bassères, D.S., Zheng, R., Zhang, P., Dayaram, T., Yamamoto, Y., Sternberg, D.W., Lokker, N., Giese, N.A., Bohlander, S.K., et al. (2006). Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. J. Exp. Med. *203*, 371–381.
78. Bolouri, H., Farrar, J.E., Triche, T., Jr., Ries, R.E., Lim, E.L., Alonzo, T.A., Ma, Y., Moore, R., Mungall, A.J., Marra, M.A., et al. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat. Med. *24*, 103–112.
79. Goossens, S., Wang, J., Tremblay, C.S., De Medts, J., T'Sas, S., Nguyen, T., Saw, J., Haigh, K., Curtis, D.J., Van Vlierberghe, P., et al. (2019). ZEB2 and LMO2 drive immature T-cell lymphoblastic leukemia via distinct oncogenic mechanisms. Haematologica *104*, 1608–1616.
80. Vuckovic, D., Bao, E.L., Akbari, P., Lareau, C.A., Mousas, A., Jiang, T., Chen, M.H., Raffield, L.M., Tardaguila, M., Huffman, J.E., et al.; VA Million Veteran Program (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell *182*, 1214–1231.e11.
81. Murakami, N., Okuno, Y., Yoshida, K., Shiraishi, Y., Nagae, G., Suzuki, K., Narita, A., Sakaguchi, H., Kawashima, N., Wang, X., et al. (2018). Integrated molecular profiling of juvenile myelomonocytic leukemia. Blood *131*, 1576–1586.
82. Osumi, T., Tsujimoto, S.I., Tamura, M., Uchiyama, M., Nakabayashi, K., Okamura, K., Yoshida, M., Tomizawa, D., Watanabe, A., Takahashi, H., et al. (2018). Recurrent *RARB* Translocations in Acute Promyelocytic Leukemia Lacking *RARA* Translocation. Cancer Res. *78*, 4452–4458.
83. Campregher, P.V., Halley, N.D.S., Vieira, G.A., Fernandes, J.F., Velloso, E.D.R.P., Ali, S., Mughal, T., Miller, V., Mangueira, C.L.P., Odone, V., and Hamerschlak, N. (2017). Identification of a novel fusion TBL1XR1-PDGFRB in a patient with acute myeloid leukemia harboring the DEK-NUP214 fusion and clinical response to dasatinib. Leuk. Lymphoma *58*, 2969–2972.