

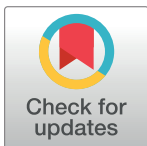
## RESEARCH ARTICLE

Evidence that 6q25.1 variant rs6931104 confers susceptibility to chronic myeloid leukemia through *RMND1* regulationYoung Min Woo<sup>1</sup>, Sehwa Kim<sup>1</sup>, Jong-Ho Park<sup>1</sup>, Nan Young Lee<sup>2</sup>, Jong-Won Kim<sup>1,3,4\*</sup>, Dennis Dong Hwan Kim<sup>5</sup>

**1** Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Korea, **2** Department of Laboratory Medicine, Kyungpook National University Chilgok Hospital, Daegu, Korea, **3** Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Samsung Medical Center, Seoul, Korea, **4** Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea, **5** Department of Medical Oncology & Hematology, Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Canada

☞ These authors contributed equally to this work.

\* [kimjw@skku.edu](mailto:kimjw@skku.edu)



## OPEN ACCESS

**Citation:** Woo YM, Kim S, Park J-H, Lee NY, Kim J-W, Kim DDH (2019) Evidence that 6q25.1 variant rs6931104 confers susceptibility to chronic myeloid leukemia through *RMND1* regulation. PLoS ONE 14(6): e0218968. <https://doi.org/10.1371/journal.pone.0218968>

**Editor:** Xiaoyan Jiang, University of British Columbia, CANADA

**Received:** March 27, 2019

**Accepted:** June 12, 2019

**Published:** June 25, 2019

**Copyright:** © 2019 Woo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2019R1A2C2002177).

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Chronic myeloid leukemia (CML) is a clonal myeloproliferative disorder. Our previous study reported novel loci as genetic markers associated with increased susceptibility to CML. The present study conducted an expression quantitative trait loci (eQTL) analysis to confirm that the single nucleotide polymorphisms (SNPs) at these loci affect the expression of candidate CML-susceptible genes. We identified that three SNPs (rs963193, rs6931104, and rs9371517) were related to the gene expression pattern of *RMND1* (Required For Meiotic Nuclear Division 1 Homolog) in both granulocytes and mononuclear cells from 83 healthy donors. Furthermore, reduced expression of *RMND1* expression was noted in CML patients compared with that in healthy individuals. We used the eQTL browsing tool to assess the regulatory information on the three associated significant SNPs, out of which rs6931104 showed strong evidence of regulatory effects. Chromatin immunoprecipitation (ChIP) assays demonstrated that A alleles of rs6931104 could significantly change the binding affinity of transcription factor (TF) RFX3 compared to the G alleles. Then, we performed *in vitro* experiments on *BCR-ABL1*-positive (*BCR-ABL1*<sup>+</sup>) cell lines. We found that expression of the CML-susceptible gene *RMND1* is affected by the binding affinity of TF RFX3, suggesting that RFX3 plays a role in *RMND1* expression. Our findings suggest potential target genes for associations of genetic susceptibility risk loci and provide further insights into the pathogenesis and mechanism of CML.

## Introduction

Chronic myeloid leukemia (CML) is a clonal disorder in hematopoietic stem cells characterized by the presence of an oncogenic *BCR-ABL1* fusion gene that encodes a constitutively

activated fusion protein carrying tyrosine kinase activity. To date, genome-wide association studies (GWASs) have revealed that single nucleotide polymorphisms (SNPs) are associated with CML development, progression, and treatment outcome [1–4]. However, beyond the established statistical associations, the functional relevance or underlying mechanisms have rarely been explored.

In our previous study, we performed a genome-wide analysis and reported novel loci associated with susceptibility to CML. In particular, the SNPs at the 6q25.1 locus were associated with CML in cohorts of both Korean and European descent, while the SNPs at the 17p11.1 locus were associated with CML in the Korean cohort only. In the present study, to understand the functional role of these SNPs, we conducted expression quantitative trait loci (eQTL) and functional analysis of related transcription factor in confirming whether the SNPs in these loci affect the expression of CML susceptibility candidate genes.

## Materials and methods

### Study design and description of participants

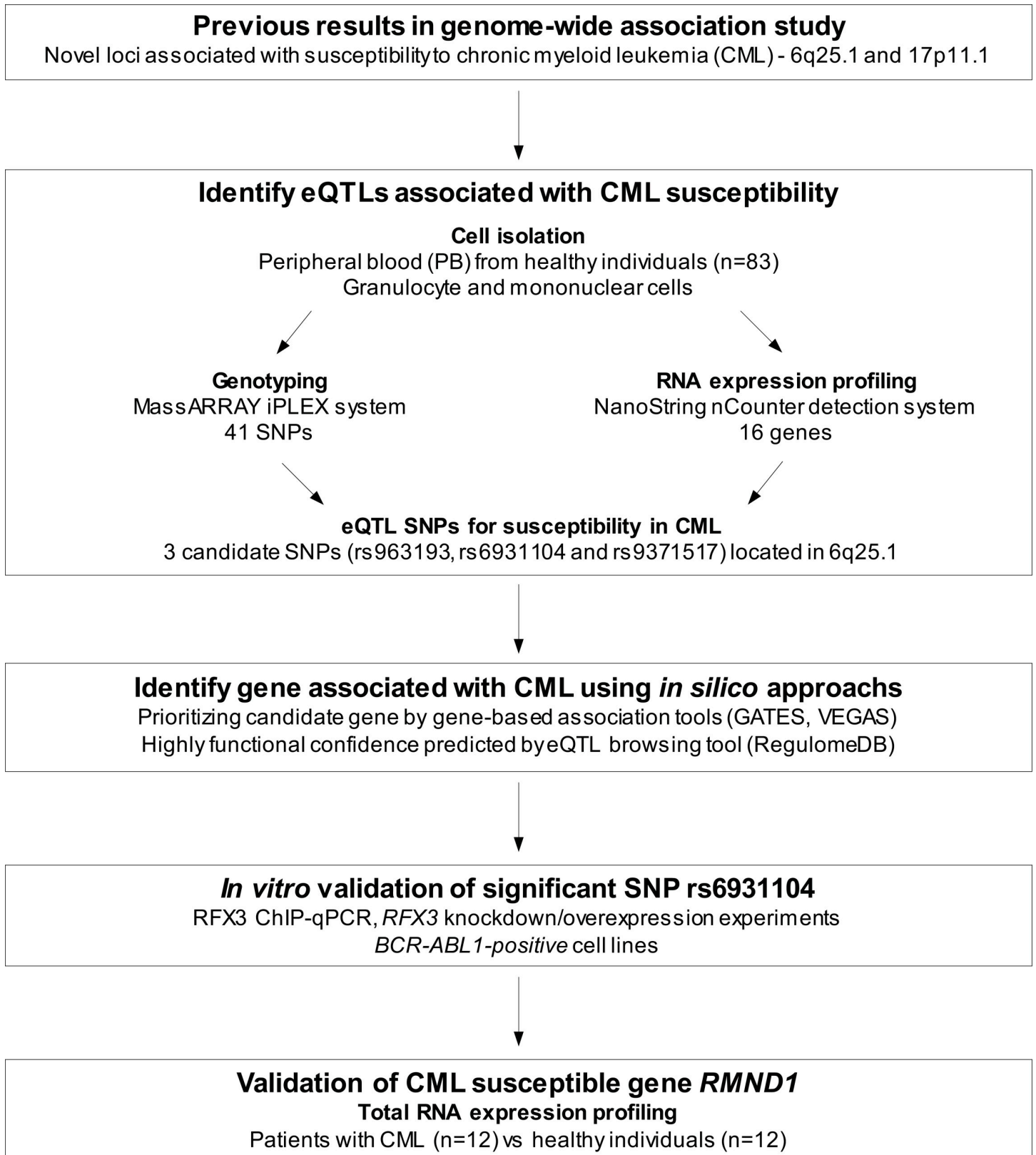
The study was approved by the Institutional Research Board of the Samsung Medical Center, Seoul, Korea. In the initial experiments, we have applied an eQTL approach in both granulocytes and mononuclear cell fraction separately to evaluate different gene expression according to the cell subsets. A total of 83 healthy donors (24 male and 59 female) who do not have any previous history of blood disorders and normal blood counts were consented and collected from March to September 2012 for this study. Granulocytes and mononuclear cells collected from peripheral blood (PB) were isolated using the Ficoll density gradient method. DNA and RNA were extracted within 24 hours after venipuncture. In order to compare the expression level of a candidate gene with healthy individuals, we additionally recruited 12 patients with CML on imatinib treatment consented for future research from March to April 2013 (Fig 1).

### Genotyping

Before evaluating whether the SNPs identified in the GWAS were functional with respect to gene expression level, we determined the genotypic variation in whole blood samples from 83 healthy individuals. Yields of pure, double-stranded genomic DNA were obtained using the QIAamp DNA Blood Maxi Kit (Qiagen, California, USA). All steps were followed according to the manufacturer's instructions. The genomic DNA of the 83 samples was quantified by a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) and diluted to 10 ng/μl. We selected 41 SNPs with *p*-values less than 0.005 selected from the discovery set in our previous study for genotyping in two loci 6q25.1 and 17p11.1, respectively [5], and added 2 SNPs (rs3736175 and rs721348) known as expression quantitative nucleotides (eQTNs, <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>), all from these regions (Table 1). Genotyping of the 41 SNPs was performed on a MassARRAY iPLEX system (Agena Bioscience, California, USA). The MassARRAY Assay Design 3.1 software was used to design multiplex primers for each SNP.

### Gene-based association analysis

We applied a gene-based association approach to prioritize susceptible genes in 6p25.1 locus associated with CML as combining the single SNP-based *p*-values using the discovery set (201 CML patients and 397 healthy individuals). We also validated the results of the gene-based approach in a validation set (356 CML patients and 1,000 healthy individuals) with Korean ethnicity.



**Fig 1. Overall study design and workflow.** To identify the eQTLs associated with CML susceptibility, we performed genotyping, RNA expression profiling, *in silico* analysis, and functional validation using healthy individuals, CML patients, and *BCR-ABL1*-positive human cell line models.

<https://doi.org/10.1371/journal.pone.0218968.g001>

Table 1. 41 SNPs selected for genotyping.

Chromosome	rs ID	Position <sup>a</sup>	p-value <sup>b</sup>
6q25.1	rs963193	151789355	4.32E-06
6q25.1	rs6931104	151786177	5.53E-06
6q25.1	rs9371517	151739810	4.05E-05
6q25.1	rs4869742	151907748	0.000415
6q25.1	rs3900024	151706570	0.0004525
6q25.1	rs7765741	151717302	0.0008366
6q25.1	rs6557140	151758042	0.002534
6q25.1	rs2295041	152773194	0.00293
6q25.1	rs9371600	152790533	0.003167
6q25.1	rs4391273	151893073	0.004039
6q25.1	rs9397105	152772494	0.00412
6q25.1	rs9397510	152797065	0.004507
6q25.1	rs7747960	152791474	0.004618
6q25.1	rs7751588	152797353	0.004929
17p11.1	rs2061893	25396146	6.47E-06
17p11.1	rs12950376	25541572	6.47E-06
17p11.1	rs1824900	25472214	6.82E-06
17p11.1	rs11080053	25388276	6.92E-06
17p11.1	rs10163524	22108783	7.09E-06
17p11.1	rs1828999	22217883	8.02E-06
17p11.1	rs7221571	25363654	8.18E-06
17p11.1	rs16972661	25398037	8.21E-06
17p11.1	rs1975980	25395691	9.70E-06
17p11.1	rs7213631	22120732	1.21E-05
17p11.1	rs1824901	25472565	1.25E-05
17p11.1	rs33962847	22084950	1.27E-05
17p11.1	rs1402724	25409526	1.37E-05
17p11.1	rs2313050	25346692	1.79E-05
17p11.1	rs11080056	25404049	2.42E-05
17p11.1	rs9972902	25455185	2.42E-05
17p11.1	rs7210668	22168795	2.51E-05
17p11.1	rs4795519	25541278	2.87E-05
17p11.1	rs2013347	22171190	3.44E-05
17p11.1	rs2340408	22235650	3.50E-05
17p11.1	rs2169956	25470744	3.78E-05
17p11.1	rs1520020	22105110	4.35E-05
17p11.1	rs12946383	21984361	0.0007233
17p11.1	rs8072903	21991614	0.0008935
17p11.1	rs7406339	22002556	0.001123
Chromosome	rs ID	Position <sup>a</sup>	Related gene
6q25.1	rs3736175	151773164	<i>RMND1</i>
17p11.1	rs7213148	25639046	<i>WSB1</i>

39 SNPs associated with CML susceptibility [5] and 2 SNPs (rs3736175 and rs721348) known as expression quantitative nucleotides (eQTNs) from these regions were selected.

<sup>a</sup>Physical position based on the human reference genome build hg19 (GRCh37).

<sup>b</sup>p-value of the discovery set in the previous study [5]

<https://doi.org/10.1371/journal.pone.0218968.t001>

The gene-based association test using extended Simes procedure (GATES) [6] is a rapid and powerful gene-based test to evaluate a gene with one or a few independent causal variants. We used the hg19 gene version to assign imputed SNPs. We built an analysis genome by rsID to optimize the annotations in the GATES dataset. The GATES tool considers that the pairwise SNP  $p$ -value correlation coefficient  $\rho$  is expected to be mainly determined by the pairwise linkage disequilibrium (LD) between the two corresponding SNPs as measured by the allelic correlation coefficient. We used HapMap JPT (Japanese) LD information data to calculate the correlation coefficient of the two SNPs. The SNPs were assigned to an extended gene region that was  $\pm 50$  kilobases (kb) from the 5' and 3' untranslated region (UTR).

The versatile gene-based association study (VEGAS) [7] method, which uses simulations from the multivariate normal distribution based on the LD structure, is a gene-level association test using the results from single SNP-based approach data [8]. We used the VEGAS software to test the 17,542 assigned genes with imputed SNPs by using the initial CML GWAS result data. To assign the SNPs to the genes, we defined the gene boundary within  $\pm 50$  kb from the 5' and 3' UTR, the same condition as for GATES. To consider the LD structure, we used HapMap JPT+CHB (Han Chinese) LD information data.

### Replication study for the gene-based approach at the 6q25.1 locus

The candidate genes identified at the 6q25.1 locus via gene-based approaches were tested in a new independent case-control replication set. To identify significant candidate genes, we selected those genes (i.e., *C6orf211*, *RMND1*, and *ZBTB2*) which had  $p < 0.001$  in the discovery set. To validate the association with CML in the replication set, we selected five observed SNPs (rs6931104, rs9371517, rs3900024, rs7765741, and rs6557140) that had  $p < 0.005$  at the 6q25.1 locus of the discovery set. The five SNPs were genotyped by the MassARRAY iPLEX system (Agena Bioscience) in the validation set ( $n = 1,356$ ).

### mRNA expression profiling

For mRNA expression profiling, total RNA was extracted from each sample of granulocyte and mononuclear cell fractions using the TRIzol reagent (Invitrogen, California, USA). A total of 16 genes were tested including 9 genes located on the 6q25.1 and 17p11.1 loci (6q25.1: *MTHFD1L*, *AKAP12*, *ZBTB2*, *RMND1*, *C6orf211*, *C6orf97*, *ESR1*; 17p11.1: *FAM27L*, *WSB1*), as well as additional 7 genes (6q25.1: *IYD*, *PLEKHG1*, *SYNE1*; 17p11.1: *KCNJ12*, *KSR1*, *LGALS9*, *NOS2A*) located in the expanded region of these loci up to 500 kb from the target genes. Also, three internal reference genes, *GAPDH*, *OAZ1*, and *PPIA*, were added. Their mRNA expression levels were quantified using the NanoString nCounter amplification-free detection system [9]. All steps were followed according to the manufacturer's instructions. Normalization of the mRNA expression was conducted by the nSolver software provided by NanoString Technologies.

### Functional prediction for eQTL SNPs

We used the RegulomeDB (<http://regulome.stanford.edu>) eQTL browsing tool to assess the regulatory information regarding the associated three significant eQTL SNPs at 6q25.1. The RegulomeDB score, which ranges from 1 to 6, is set up to assess the degree of confidence supporting a functional variant as a transcription factor (TF) binding site. A lower score indicates stronger evidence that an SNP affects binding and gene expression [10].

## Chromatin immunoprecipitation

We performed chromatin immunoprecipitation (ChIP) in 18 healthy individuals who did not have any previous history of blood disorder with normal blood counts. ChIP assays were performed using a Pierce Agarose ChIP kit (Thermo Fisher Scientific, Illinois, USA), and chromatin was cross-linked, fragmented, immunoprecipitated, and purified according to the manufacturer instructions using a rabbit polyclonal antibody against RFX3 (T-17, sc-10662; Santa Cruz Biotechnology, California, USA) or an anti-RNA polymerase II antibody (positive)/normal rabbit IgG (negative) as control (Thermo Fisher Scientific, Illinois, USA). Quantitative reverse transcription polymerase chain reaction (RT-qPCR; Applied Biosystems 7900 HT Real-Time PCR) was used to determine the ChIP assay results.

## Cell culture for *in vitro* assay

We used *BCR-ABL1*<sup>+</sup> cell lines, including K562 (CCL-243; obtained from the American Type Culture Collection), CML-T1, and LAMA84 (ACC-7 and ACC-168; obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH). All cell lines were cultured in RPMI-1640 media (Invitrogen, California, USA) supplemented with 10% fetal bovine serum (FBS; Invitrogen, California, USA), 100 units/mL penicillin, and 100 µg/mL streptomycin (Invitrogen, California, USA).

## Gene knockdown and overexpression experiments

To further explore the link between *RMND1* gene expression and TF RFX3 binding affinity, we performed *RFX3* gene knockdown and overexpression experiments. K562, CML-T1, and LAMA84 *BCR-ABL1*<sup>+</sup> cells were transfected with *RFX3* siRNA (L-011764-00-0005; ON-TARGETplus, SMARTpool, Dharmacon, Colorado, USA) and *RFX3* overexpression vector (RC201137; True ORF Gold cDNA Clones, Origene, Maryland, USA). Non-silencing siRNA was also used as a negative control (51-01-14-04; Universal Negative Control DsiRNA, IDT, Iowa, USA). Each was transfected into the cells by electroporation using the Invitrogen Neon Transfection System as previously described [11]. The cells were then resuspended in medium containing 20% FBS and further incubated at 37°C. We performed RT-qPCR to evaluate *RFX3* gene knockdown or overexpression one day after transfection. We performed more than three independent experiments with at least triplicates to ensure accuracy.

## Gene expression analysis in CML patients and healthy individuals

To investigate the relationship between *RMND1* expression and CML, the *RMND1* mRNA level was assessed by RT-qPCR in CML patients and healthy individuals. Total cDNA was amplified using the SYBR Select Master Mix (Thermo Fisher Scientific, Texas, USA). After 45 cycles at 95°C for 5 sec and 60°C for 1 min, the dissociation curve analysis for  $\Delta$ Ct was analyzed using the SDS 2.4 and RQ Manager software provided by Life Technologies. For quantification, the RNA transcript expression was normalized by determining the ratio between the expression levels of the *RMND1* and *GAPDH* genes.

## Statistical analysis

We carried out an imputation analysis to fill in the missing genotype data using IMPUTE2 Version 2.3.0 [12, 13]. As a reference panel, the JPT + CHB (Phase II HapMap release #22 NCBI build 36) populations were used [14]. We also used the pre-phasing method SHAPEIT2 version v2.r727 for efficient imputation with a small loss in accuracy [8, 15]. The LD structure was assessed using Haploview version 4.2.7 [16]. The eQTL analysis was carried out using the

ANOVA model provided by the Matrix eQTL software [17]. We applied the family-wise error (or  $p$ -value) correction using the False Discovery Rate (FDR) method. All statistical analyses were performed using R (R Foundation for Statistical Computing, Austria).

## Results

### *RMND1* expression is correlated with three SNP genotypes at 6q25.1

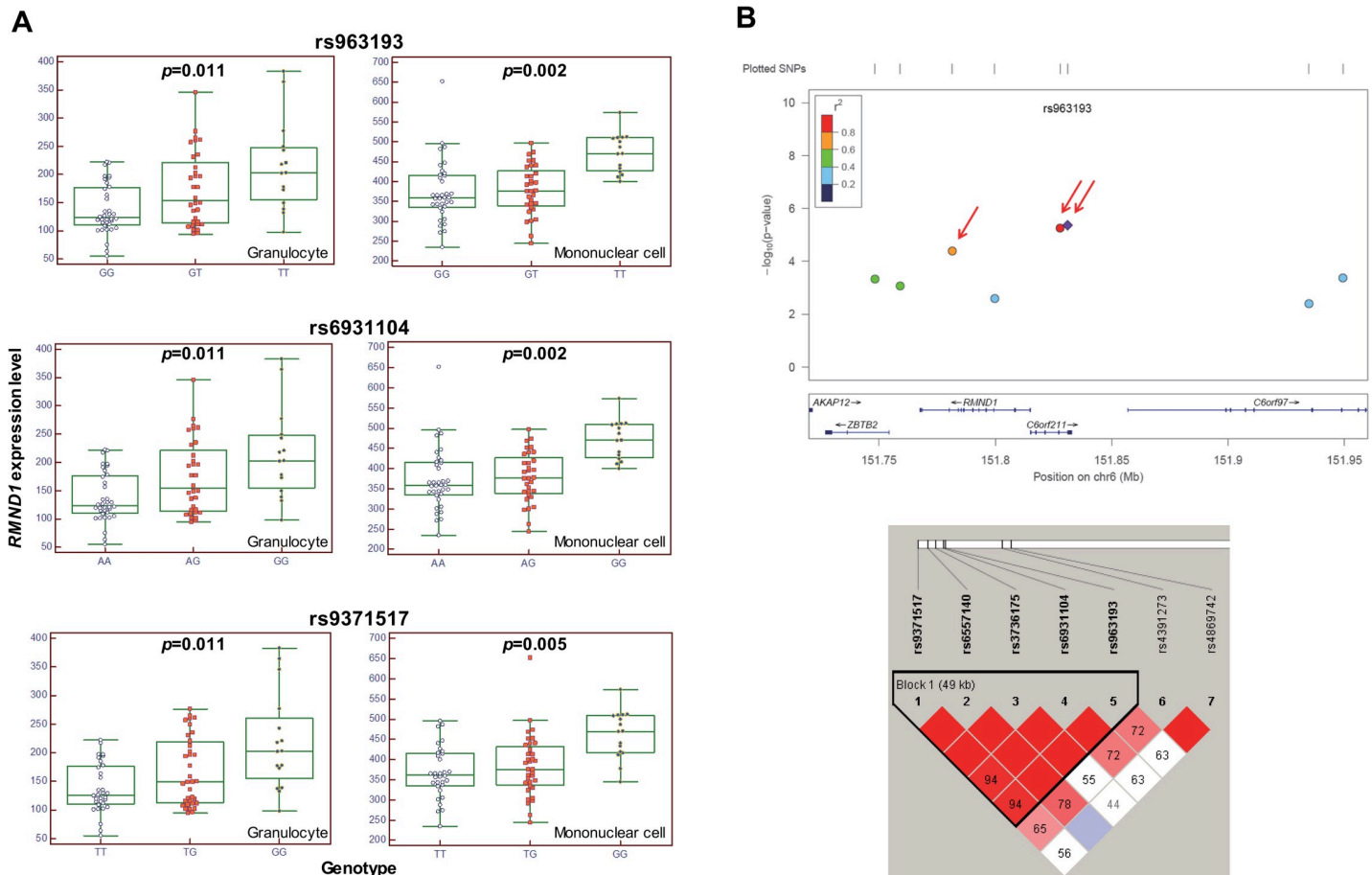
We examined the associations of 41 SNPs at two loci, 6q25.1 and 17p11.1 (Table 1), in 83 healthy individuals to evaluate whether those SNPs were functional. We focused on nine genes that had previously been reported [5] in the regions of 6q25.1 (*MTHFD1L*, *AKAP12*, *ZBTB2*, *RMND1*, *C6orf211*, *C6orf97*, *ESR1*) and 17p11.1 (*FAM27L*, *WSB1*). Additional seven genes located up to 500 kb from the target genes were added (6q25.1: *IYD*, *PLEKHG1*, *SYNE1*; 17p11.1: *KCNJ12*, *KSR1*, *LGALS9*, *NOS2A*) to investigate any possible regulatory functions on the expression of neighboring genes. We found evidence that 3 candidate SNPs (rs963193, rs6931104, and rs9371517) located in non-coding regions of 6q25.1 were related to the gene expression pattern of *RMND1* in both granulocytes (corrected  $p$ : 0.011, 0.011, and 0.011, respectively) and mononuclear cells (corrected  $p$ : 0.002, 0.002, and 0.005, respectively). The genotype-dependent mRNA expression levels of *RMND1* showed similar patterns for rs963193 and rs6931104 (Fig 2A). A subset of significant SNPs identified as eQTLs were in strong LD, with an  $r^2$  value = 1. The LD block shows that the three SNPs are in the same block (Fig 2B). However, one SNP allocated to 17p11.1 was not found to be associated with the expression level of candidate genes at this locus.

### *RMND1* is significantly associated with CML in the two gene-based association tools, VEGAS and GATES

Using VEGAS, we identified three top-ranking genes (*RMND1*,  $p = 0.000011$ ; *C6orf211*,  $p = 0.000016$ ; and *ZBTB2*,  $p = 0.00029$ ) in the candidate locus (6q25.1) associated with CML. We also confirmed the three genes associated with CML by another gene-based method, GATES, at the threshold  $p$ -value less than 0.001. Notably, *RMND1* ( $p = 0.011$  for VEGAS and  $p = 0.0088$  for GATES) and *ZBTB2* ( $p = 0.0093$  for VEGAS and  $p = 0.042$  for GATES) were validated in an independent cohort of Korean subjects (total  $n = 1,356$ ; case  $n = 356$  and control  $n = 1,000$ ) with more than one SNP with  $p < 0.005$ , which constituted original genotype, rather than imputed data. It is important to improve the probability that specific genes are on loci associated with CML because functional studies are not feasible for all located genes. In conclusion, our gene-based approach has identified the susceptible genes associated with CML.

### Non-coding variant rs6931104 has the strongest evidence of regulatory potential

We further investigated the regulatory effects (i.e., transcription factor binding sites, chromatin peak regions, and DNase accessibility) of the three eQTL SNPs (rs963193, rs6931104, and rs9371517) using RegulomeDB. Remarkably, rs6931104 (score = 1) is in LD with the significant  $p$ -value eQTLs rs963193 and rs9371517, SNPs that by themselves are not functional according to RegulomeDB (score = 6). rs6931104, the SNP with strong evidence of a regulatory function, is an intergenic SNP located in the *C6orf211* region and 12.9 kb upstream from the transcription start site (TSS) of *RMND1*. Specifically, RegulomeDB annotated rs6931104 as a putative binding site for the transcription factor (TF) of RFX3 (Regulatory Factor X3) (S1 Fig).



**Fig 2. *RMND1* expression is correlated with the genotype of the functional SNP rs6931104 at 6q25.1.** (A) Box-plot of the significant eQTLs in both cells. *RMND1* expression levels in granulocytes (left) and mononuclear cells (right) (corrected  $p < 0.05$ ). (B) Manhattan plot and LD block of 6q25.1, including three significant SNPs (rs9371517, rs6931104, and rs963193). The Manhattan plot shows the  $p$ -value of the discovery set in the previous study, and the red arrows indicate these three SNPs. The LD block image was created based on the genotype information in this study.

<https://doi.org/10.1371/journal.pone.0218968.g002>

Therefore, rs6931104 may be involved in the regulation of *RMND1* mRNA expression by regulating TF RFX3 binding.

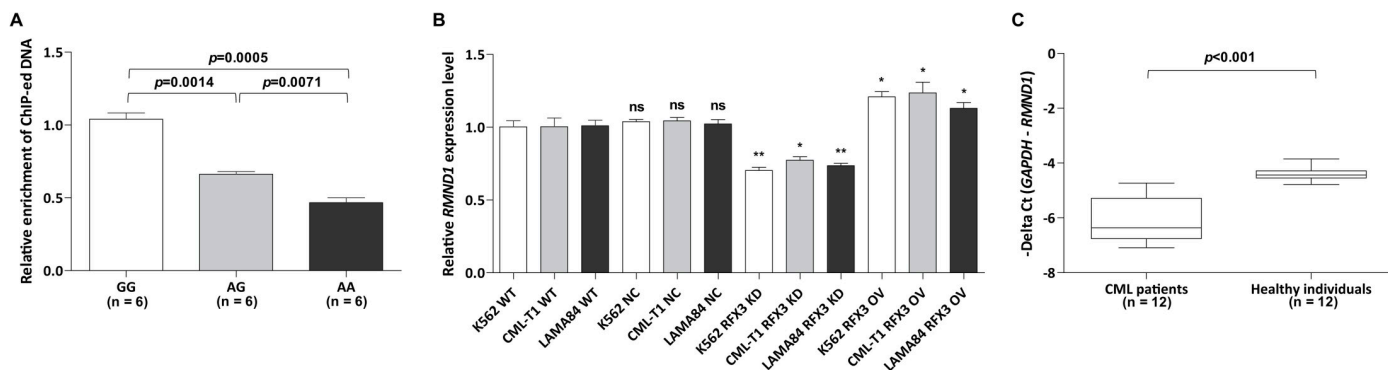
### Change in RFX3 binding affinity is correlated with change in *RMND1* expression

To confirm this hypothesis, we performed ChIP-qPCR on 18 healthy individuals. As expected, the functional assays demonstrated that the presence of the A allele of the functional SNP rs6931104 showed the decreased binding affinity of the RFX3 transcription factor compared to the presence of the G allele. The amount of immunoprecipitated DNA was lower for either the AG or AA genotype compared to that for the GG genotype (Fig 3A). Thus, we postulated that the *RMND1* gene expression is dependent on the binding affinity of TF RFX3.

### RFX3 abundance is reflected in a change in *RMND1* expression in cell line models

To assess whether there was a correlation between *RMND1* gene expression and TF RFX3 abundance, we performed RFX3 gene knockdown and overexpression in *BCR-ABL1*<sup>+</sup> cell





**Fig 3. Expression of CML-susceptible gene *RMND1* is affected by TF RFX3 binding affinity.** (A) The CML-associated SNP rs6931104 affects the binding affinity of transcription factor RFX3. ChIP-ed DNA enrichment determined by qRT-PCR is graphed as the amount of immunoprecipitated DNA. Error bars indicate SD values. (B) Effects of *RFX3* knockdown (KD) or overexpression (OV) on *RMND1* transcript levels in *BCR-ABL1*<sup>+</sup> cells. Total RNA expression was assayed by RT-qPCR. Asterisk marked *p*-values were calculated by comparing with control in each of the three cell lines, respectively. \*\**p* < 0.01, \**p* < 0.05, ns—not significant. Error bars indicate SD values. (C) *RMND1* transcript levels in CML patients. Total RNA expression was assayed by RT-qPCR. Minus delta Ct (*GAPDH* Ct value–*RMND1* Ct value) indicates the expression levels (Y-axis) in samples from 12 human CML patients and 12 healthy individuals. All statistical significance was evaluated by the Student's *t*-test.

<https://doi.org/10.1371/journal.pone.0218968.g003>

lines. Using quantitative RT-PCR, *RFX3* knockdown was found to reduce *RMND1* mRNA levels compared to the respective control significantly. In contrast, *RFX3* overexpression resulted in a significant increase in *RMND1* mRNA abundance (Fig 3B). Using quantitative RT-PCR in the *BCR-ABL1*<sup>+</sup> cells (K562, CML-T1, and LAMA84), we verified the decreased expression of *RMND1* with approximately 30%, 23%, and 27% compared to control, respectively. In *RFX3*-overexpressed cells, the quantitative RT-PCR results verified that *RMND1* expression levels were significantly increased by approximately 24%, 27%, and 16% in the stable *RFX3*-overexpressed cells compared to the control cells, respectively. These data suggest that transcriptional regulation by *RFX3* is likely to affect *RMND1* expression.

### *RMND1* is downregulated in CML patients

*RMND1* expression levels were determined by RT-qPCR in samples from 12 human CML patients and 12 healthy individuals. As shown in Fig 3C, *RMND1* expression was significantly downregulated in CML patients compared to that in healthy individuals (*p* < 0.001). This finding on the relationship between the low *RMND1* levels and the CML samples supports the tumor-related role of *RMND1* in CML.

## Discussion

The eQTL study has emerged as an essential concept to investigate the biological effects of non-coding regulatory variants [18]. According to the results of the genotype-tissue expression (GTEx) project, nearly 50% of the genetic variants associated with human disease colocalize with an eQTL [19]. This means that eQTL analysis provides insights enabling a pathological understanding of non-coding variants associated with the disease.

The key finding of the present study is that the causal eQTL variant rs6931104 at 6q25.1 is capable of driving a differential expression of *RMND1* and that this action is at least in part dependent on the binding of TF RFX3. Several genes are located within a 1 Mb region of the 6q25.1 locus, including *ZBTB2*, *RMND1* [*C6orf96*], *C6orf97*, *C6orf211*, and *AKAP12* [5]. Recently, genome-wide studies have identified relationships in the expression patterns between genes and the SNPs in this region, which is upstream of the gene encoding ER (*ESR1*), which is associated with breast cancer susceptibility [20–24]. Dunbier and colleagues

have previously suggested that some of the biological effects caused by *ESR1* could be mediated and modified by these co-expressed genes [21].

The CML-susceptible gene *RMND1* encodes a protein composed of 449 amino acids that localize to the mitochondria and is shown to behave as an integral membrane protein [25]. It belongs to the evolutionarily conserved proteins that share the DUF155 domain (Domain of Unknown Function 155) [26, 27]. However, its exact function remains unknown. Several studies have suggested that the *RMND1* gene is associated with human diseases, with lower *RMND1* expression correlating with breast cancer risk and worse relapse-free survival [28, 29], infantile encephalopathy, mitochondrial translation defects, hearing impairment, and renal failure [26, 30]. Moreover, *RMND1* orthologs are involved in cell division [31], and mitochondrial ribosomal proteins affect cell-cycle regulation [25, 32], suggesting a possibility that *RMND1* gene expression is associated with the proliferation of mature granulocytes and their precursors, a major characteristic of CML.

Non-coding SNPs that show associations with gene expression levels can constitute strong clues for the cause of certain diseases, such as SNPs located in transcription factor binding sites (TFBS) that can affect gene regulation by altering the binding affinity of the respective TFs [33, 34]. We demonstrated that the risk-associated SNP rs6931104 could significantly change the binding affinity of transcription factor RFX3. RFX3, also known as Regulatory Factor X3, has been characterized as a transcriptional activator that can bind DNA with other RFX family members [35, 36]. The biological relevance of RFX3 in disease has never been extensively investigated, and it is not known whether RFX3 is involved in tumorigenesis, particularly in CML. Thus, further functional studies of RFX3 are needed to elucidate the role of RFX3 in CML leukemogenesis based on the results of the current study.

In conclusion, we conducted an eQTL analysis and functional validation for lead variants and nearby genes using data from a genome-wide analysis on a large number of CML patients. The risk SNP rs6931104 lowers the binding affinity of TF RFX3; it also reduces the expression of *RMND1*. These results suggest that a reduction in the expression levels of *RMND1*, which serves as a tumor suppressor gene, is associated with CML pathogenesis. Theoretically, *RMND1* regulation can be a potential therapeutic strategy to improve treatment outcome in CML patients. In the present study, we focused on whether SNPs at a risk locus affected candidate gene expression levels using transcriptional eQTL. However, our results imply that other regulatory mechanisms may also be involved in the pathogenesis of CML. They also suggest that various types of modulation (i.e., chromatin modification QTL [37], chromatin interaction QTL [38], and translational QTL [39]) may affect CML pathogenesis. Further functional studies are needed to elucidate the role of other genes in CML leukemogenesis based on the results of the current study.

In our study, we have discovered additional biological mechanisms for understanding genetic susceptibility risk loci, and have provided a more profound insight into the genetic and biological basis for the pathogenesis of CML.

## Supporting information

**S1 Fig. rs6931104 showed strong evidence of regulatory effects.** Functional information of rs6931104 (<http://regulome.stanford.edu>). SNPs categorized as '1f' were described to function as eQTLs and be related with either TF binding or DNase, which means that they have high functional confidence. (PDF)

**S2 Fig. Effects of RFX3 knockdown (KD) or overexpression (OV) on RFX3 transcript levels in *BCR-ABL1*<sup>+</sup> cells.** Relative transcript levels of *RFX3* were quantified by RT-qPCR and

normalized with a house-keeping gene *GAPDH*. The results show are representative of at least three independent experiments. Statistical analysis was performed using Student's *t*-test.

\*\*\* $p < 0.001$ , ns—not significant. Error bars indicate SD values.

(PDF)

**S1 Table. Primer sequences used for *in vitro* assay.**

(XLSX)

## Author Contributions

**Conceptualization:** Jong-Won Kim, Dennis Dong Hwan Kim.

**Data curation:** Jong-Ho Park.

**Formal analysis:** Young Min Woo, Sehwa Kim, Jong-Ho Park.

**Funding acquisition:** Jong-Won Kim.

**Investigation:** Young Min Woo, Sehwa Kim, Jong-Ho Park.

**Methodology:** Young Min Woo, Sehwa Kim, Jong-Ho Park.

**Project administration:** Jong-Won Kim.

**Supervision:** Jong-Won Kim, Dennis Dong Hwan Kim.

**Writing – original draft:** Young Min Woo, Sehwa Kim, Jong-Ho Park.

**Writing – review & editing:** Nan Young Lee, Jong-Won Kim, Dennis Dong Hwan Kim.

## References

1. Crossman LC, O'Hare T, Lange T, Willis SG, Stoffregen EP, Corbin AS, et al. A single nucleotide polymorphism in the coding region of ABL and its effects on sensitivity to imatinib. *Leukemia*. 2005; 19(11):1859–62. Epub 2005/09/10. <https://doi.org/10.1038/sj.leu.2403935> PMID: 16151465.
2. Dulucq S, Bouchet S, Turcq B, Lippert E, Etienne G, Reiffers J, et al. Multidrug resistance gene (MDR1) polymorphisms are associated with major molecular responses to standard-dose imatinib in chronic myeloid leukemia. *Blood*. 2008; 112(5):2024–7. Epub 2008/06/06. <https://doi.org/10.1182/blood-2008-03-147744> PMID: 18524988.
3. Kim DH, Xu W, Ma C, Liu X, Siminovitch K, Messner HA, et al. Genetic variants in the candidate genes of the apoptosis pathway and susceptibility to chronic myeloid leukemia. *Blood*. 2009; 113(11):2517–25. Epub 2009/01/15. <https://doi.org/10.1182/blood-2008-07-169110> PMID: 19141860.
4. Boulwood J, Perry J, Zaman R, Fernandez-Santamaria C, Littlewood T, Kusec R, et al. High-density single nucleotide polymorphism array analysis and ASXL1 gene mutation screening in chronic myeloid leukemia during disease progression. *Leukemia*. 2010; 24(6):1139–45. Epub 2010/04/23. <https://doi.org/10.1038/leu.2010.65> PMID: 20410925.
5. Kim DH, Lee ST, Won HH, Kim S, Kim MJ, Kim HJ, et al. A genome-wide association study identifies novel loci associated with susceptibility to chronic myeloid leukemia. *Blood*. 2011; 117(25):6906–11. Epub 2011/05/05. <https://doi.org/10.1182/blood-2011-01-329797> PMID: 21540461.
6. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet*. 2011; 88(3):283–93. Epub 2011/03/15. <https://doi.org/10.1016/j.ajhg.2011.01.019> PMID: 21397060; PubMed Central PMCID: PMC3059433.
7. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010; 87(1):139–45. Epub 2010/07/06. <https://doi.org/10.1016/j.ajhg.2010.06.009> PMID: 20598278; PubMed Central PMCID: PMC2896770.
8. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013; 10(1):5–6. Epub 2012/12/28. <https://doi.org/10.1038/nmeth.2307> PMID: 23269371.
9. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Curr Protoc Mol Biol*. 2011; Chapter 25:Unit25B 10. Epub 2011/04/08. <https://doi.org/10.1002/0471142727.mb25b10s94> PMID: 21472696.

10. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22(9):1790–7. Epub 2012/09/08. <https://doi.org/10.1101/gr.137323.112> PMID: 22955989; PubMed Central PMCID: PMC3431494.
11. Park JH, Woo YM, Youm EM, Hamad N, Won HH, Naka K, et al. HMGCLL1 is a predictive biomarker for deep molecular response to imatinib therapy in chronic myeloid leukemia. *Leukemia.* 2018. Epub 2018/12/18. <https://doi.org/10.1038/s41375-018-0321-8> PMID: 30555164.
12. 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):e1000529. Epub 2009/06/23. <https://doi.org/10.1371/journal.pgen.1000529> PMID: 19543373; PubMed Central PMCID: PMC2689936.
13. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012; 44(8):955–9. Epub 2012/07/24. <https://doi.org/10.1038/ng.2354> PMID: 22820512; PubMed Central PMCID: PMC3696580.
14. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449(7164):851–61. Epub 2007/10/19. <https://doi.org/10.1038/nature06258> PMID: 17943122; PubMed Central PMCID: PMC2689609.
15. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2011; 9(2):179–81. Epub 2011/12/06. <https://doi.org/10.1038/nmeth.1785> PMID: 22138821.
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21(2):263–5. Epub 2004/08/07. <https://doi.org/10.1093/bioinformatics/bth457> PMID: 15297300.
17. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012; 28(10):1353–8. Epub 2012/04/12. <https://doi.org/10.1093/bioinformatics/bts163> PMID: 22492648; PubMed Central PMCID: PMC3348564.
18. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015; 16(4):197–212. Epub 2015/02/25. <https://doi.org/10.1038/nrg3891> PMID: 25707927.
19. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017; 550(7675):204–13. Epub 2017/10/13. <https://doi.org/10.1038/nature24277> PMID: 29022597; PubMed Central PMCID: PMC5776756.
20. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet.* 2009; 41(3):324–8. Epub 2009/02/17. <https://doi.org/10.1038/ng.318> PMID: 19219042; PubMed Central PMCID: PMC2754845.
21. Dunbier AK, Anderson H, Ghazoui Z, Lopez-Knowles E, Pancholi S, Ribas R, et al. ESR1 is co-expressed with closely adjacent uncharacterised genes spanning a breast cancer susceptibility locus at 6q25.1. *PLoS Genet.* 2011; 7(4):e1001382. Epub 2011/05/10. <https://doi.org/10.1371/journal.pgen.1001382> PMID: 21552322; PubMed Central PMCID: PMC3084198 as an expert witness for them.
22. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet.* 2010; 42(6):504–7. Epub 2010/05/11. <https://doi.org/10.1038/ng.586> PMID: 20453838; PubMed Central PMCID: PMC3632836.
23. Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet.* 2010; 6(7):e1001029. Epub 2010/07/28. <https://doi.org/10.1371/journal.pgen.1001029> PMID: 20661439; PubMed Central PMCID: PMC2908678.
24. Cai Q, Wen W, Qu S, Li G, Egan KM, Chen K, et al. Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of chinese, Japanese, and European ancestry. *Cancer Res.* 2011; 71(4):1344–55. Epub 2011/02/10. <https://doi.org/10.1158/0008-5472.CAN-10-2733> PMID: 21303983; PubMed Central PMCID: PMC3083305.
25. Garcia-Diaz B, Barros MH, Sanna-Cherchi S, Emmanuele V, Akman HO, Ferreiro-Barros CC, et al. Infantile encephalomyopathy and defective mitochondrial translation are due to a homozygous *RMND1* mutation. *Am J Hum Genet.* 2012; 91(4):729–36. Epub 2012/10/02. <https://doi.org/10.1016/j.ajhg.2012.08.019> PMID: 23022099; PubMed Central PMCID: PMC3484479.
26. Janer A, Antonicka H, Lalonde E, Nishimura T, Sasarman F, Brown GK, et al. An *RMND1* Mutation causes encephalopathy associated with multiple oxidative phosphorylation complex deficiencies and a mitochondrial translation defect. *Am J Hum Genet.* 2012; 91(4):737–43. Epub 2012/10/02. <https://doi.org/10.1016/j.ajhg.2012.08.020> PMID: 23022098; PubMed Central PMCID: PMC3484649.
27. Ng YS, Alston CL, Diodato D, Morris AA, Ulrick N, Kmoch S, et al. The clinical, biochemical and genetic features associated with *RMND1*-related mitochondrial disease. *J Med Genet.* 2016; 53(11):768–75. Epub 2016/07/15. <https://doi.org/10.1136/jmedgenet-2016-103910> PMID: 27412952; PubMed Central PMCID: PMC5264221.

28. Dunning AM, Michailidou K, Kuchenbaecker KB, Thompson D, French JD, Beesley J, et al. Breast cancer risk variants at 6q25 display different phenotype associations and regulate *ESR1*, *RMND1* and *CCDC170*. *Nat Genet.* 2016; 48(4):374–86. Epub 2016/03/02. <https://doi.org/10.1038/ng.3521> PMID: 26928228; PubMed Central PMCID: PMC4938803.
29. Yamamoto-Ibusuki M, Yamamoto Y, Fujiwara S, Sueta A, Yamamoto S, Hayashi M, et al. C6ORF97-*ESR1* breast cancer susceptibility locus: influence on progression and survival in breast cancer patients. *Eur J Hum Genet.* 2015; 23(7):949–56. Epub 2014/11/06. <https://doi.org/10.1038/ejhg.2014.219> PMID: 25370037; PubMed Central PMCID: PMC4463506.
30. Ravn K, Neland M, Wibrand F, Duno M, Ostergaard E. Hearing impairment and renal failure associated with *RMND1* mutations. *Am J Med Genet A.* 2016; 170A(1):142–7. Epub 2015/09/24. <https://doi.org/10.1002/ajmg.a.37399> PMID: 26395190.
31. Zhou X, Li Q, Chen X, Liu J, Zhang Q, Liu Y, et al. The Arabidopsis *RETARDED ROOT GROWTH* gene encodes a mitochondria-localized protein that is required for cell division in the root meristem. *Plant Physiol.* 2011; 157(4):1793–804. Epub 2011/10/11. <https://doi.org/10.1104/pp.111.185827> PMID: 21984726; PubMed Central PMCID: PMC3327206.
32. Yoo YA, Kim MJ, Park JK, Chung YM, Lee JH, Chi SG, et al. Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. *Mol Cell Biol.* 2005; 25(15):6603–16. Epub 2005/07/19. <https://doi.org/10.1128/MCB.25.15.6603-6616.2005> PMID: 16024796; PubMed Central PMCID: PMC1190350.
33. Wang X, Tomso DJ, Liu X, Bell DA. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol Appl Pharmacol.* 2005; 207(2 Suppl):84–90. Epub 2005/07/09. <https://doi.org/10.1016/j.taap.2004.09.024> PMID: 16002116.
34. Jiang J, Jia P, Shen B, Zhao Z. Top associated SNPs in prostate cancer are significantly enriched in cis-expression quantitative trait loci and at transcription factor binding sites. *Oncotarget.* 2014; 5(15):6168–77. Epub 2014/07/16. <https://doi.org/10.18632/oncotarget.2179> PMID: 25026280; PubMed Central PMCID: PMC4171620.
35. Reith W, Ucla C, Barras E, Gaud A, Durand B, Herrero-Sanchez C, et al. RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. *Mol Cell Biol.* 1994; 14(2):1230–44. Epub 1994/02/01. <https://doi.org/10.1128/mcb.14.2.1230> PMID: 8289803; PubMed Central PMCID: PMC358479.
36. Emery P, Durand B, Mach B, Reith W. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res.* 1996; 24(5):803–7. Epub 1996/03/01. <https://doi.org/10.1093/nar/24.5.803> PMID: 8600444; PubMed Central PMCID: PMC145730.
37. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell.* 2015; 162(5):1051–65. Epub 2015/08/25. <https://doi.org/10.1016/j.cell.2015.07.048> PMID: 26300125; PubMed Central PMCID: PMC4556133.
38. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell.* 2015; 163(7):1611–27. Epub 2015/12/22. <https://doi.org/10.1016/j.cell.2015.11.024> PMID: 26686651; PubMed Central PMCID: PMC4734140.
39. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Genomic variation. Impact of regulatory variation from RNA to protein. *Science.* 2015; 347(6222):664–7. Epub 2015/02/07. <https://doi.org/10.1126/science.1260793> PMID: 25657249; PubMed Central PMCID: PMC4507520.