ORIGINAL ARTICLE

Molecular Genetics & Genomic Medicine Open Access **WILEY**

Genome-wide association study of café-au-lait macule number in neurofibromatosis type 1

Heejong Sung[1] [ID] | **Paula L. Hyland[2,3]** | **Alexander Pemov[4]** [ID] | **Jeremy A. Sabourin[1]** | **Andrea M. Baldwin[5]** | **Sara Bass[6]** | **Kedest Teshome[6]** | **Wen Luo[6]** | **Frederick National Laboratory for Cancer Research[6]** | **Brigitte C. Widemann[5]** | **Douglas R. Stewart[4]** [ID] | **Alexander F. Wilson[1]**

[1]Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA

[2]Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

[3]Division of Applied Regulatory Science, Office of Translational Science, Center for Drug Evaluation & Research, U.S. Food and Drug Administration, Silver Spring, MD, USA

[4]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

[5]Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

[6]Frederick National Laboratory for Cancer Research, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

Correspondence
Douglas R. Stewart, Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA.
Email: drstewart@mail.nih.gov

Alexander F. Wilson, Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA.
Email: afw@mail.nih.gov

Funding information
NIH

Abstract

Background: Neurofibromatosis type 1 (NF1) is a tumor-predisposition disorder that arises due to pathogenic variants in tumor suppressor *NF1*. NF1 has variable expressivity that may be due, at least in part, from heritable elements such as modifier genes; however, few genetic modifiers have been identified to date.

Methods: In this study, we performed a genome-wide association analysis of the number of café-au-lait macules (CALM) that are considered a tumor-like trait as a clinical phenotype modifying NF1.

Results: A borderline genome-wide significant association was identified in the discovery cohort (CALM1, $N = 112$) between CALM number and rs12190451 (and rs3799603, $r^2 = 1.0$; $p = 7.4 \times 10^{-8}$) in the intronic region of *RPS6KA2*. Although, this association was not replicated in the second cohort (CALM2, $N = 59$) and a meta-analysis did not show significantly associated variants in this region, a significant corroboration score (0.72) was obtained for the *RPS6KA2* signal in the discovery cohort (CALM1) using Complementary Pairs Stability Selection for Genome-Wide Association Studies (ComPaSS-GWAS) analysis, suggesting that the lack of replication may be due to heterogeneity of the cohorts rather than type I error.

Heejong Sung, Paula L. Hyland, and Alexander Pemov contributed equally to this work.

Conclusion: rs12190451 is located in a melanocyte-specific enhancer and may influence *RPS6KA2* expression in melanocytes—warranting further functional studies.

KEYWORDS

café-au-lait macule, complementary pairs stability selection for genome-wide association studies analysis, genetic modifiers, genome-wide association study, neurofibromatosis type 1

1 | INTRODUCTION

Neurofibromatosis type 1 (NF1) is a common, monogenic tumor-predisposition disorder that arises due to germline pathogenic variation in *NF1* (OMIM 613113). The protein product of *NF1*, neurofibromin, regulates the conversion of the active form of RAS (RAS-GTP) to its inactive form (RAS-GDP; Welti, D'Angelo, & Scheffzek, 2008). Individuals haplo-insufficient for *NF1* have increased levels of RAS-GTP, dysregulation of cell growth, and thus have an increased risk of a variety of benign and malignant tumors, predominantly of the central and peripheral nervous system (e.g., neurofibromas, pilocytic astrocytomas, optic pathway gliomas; Huson, 2008). Soft-tissue sarcomas (malignant peripheral nerve sheath tumors, rhabdomyosarcomas), neuroendocrine tumors (somatostatinomas, pheochromocytomas), and leukemia (juvenile myelomonocytic leukemia) are also associated with NF1 (Brems, Beert, de Ravel, & Legius, 2009). Although the first, modern medical description of NF1 was published in 1882 by von Recklinghausen, novel tumor associations continue to be discovered, including gastrointestinal stromal tumors (Stewart et al., 2007), glomus tumors (Brems, Park, et al., 2009), and breast cancer (Sharif et al., 2007; Uusitalo et al., 2016).

NF1 has variable expressivity, even among members of the same family. This phenotypic complexity probably derives from multiple etiologies (e.g., epigenetic phenomena and stochastic events) as well as from heritable elements, such as modifier genes (Carey & Viskochil, 1999). There is clinical evidence that modifier genes are a major component of phenotypic variation in NF1. Easton, Ponder, Huson, and Ponder (1993) first examined intra-familial correlation in phenotypic features (Easton et al., 1993). In that study, among six pairs of monozygotic twins, the highest correlations were for the number of café-au-lait spots ($r = 0.85$, $p < .05$) and cutaneous neurofibromas ($r = 1.00$, $p < .01$). Correlation decreased for first- and second-degree relatives. Four of the five binary traits studied in that paper also showed significant familial clustering (Easton et al., 1993). Szudek, Joe, and Friedman (2002) observed similar patterns of familial correlation that also suggested a role for genetic factors (Szudek et al., 2002; Szudek et al., 2002). A statistical analysis of NF1 phenotypes and severity in a French cohort provided additional evidence for the existence of genetic modifiers in NF1 (Sabbagh et al., 2009). The heritability of café-au-lait macules, a hallmark feature of NF1, was the highest (0.62 ± 0.08) among the traits considered.

In people with NF1, CALM number is a phenotype that has a number of advantages: compared to other NF1 clinical features it is relatively stable during patients' life (although some "fading" is observed), is readily quantifiable and can be non-invasively measured. Moreover, CALMs are "tumor-like" in that they arise from a somatic inactivation (a second hit) of the remaining normal copy of *NF1* in skin melanocytes, which, like the Schwann cells, also originate from the common neural crest cell progenitor (Maertens et al., 2007). In previous work using a gene expression-based, multi-platform approach, variants were identified in the intergenic region of *DPH2* (OMIM 603456) and *ATP6V0B* (OMIM 603717) and in *MSH6* (OMIM 600678) that influence CALM number in NF1 (Pemov et al., 2014). In the current work, a genome-wide association study (GWAS) was employed for the first time to identify additional variants associated with CALM number in NF1. Furthermore, the potential functional effects of these variants were evaluated in both normal skin and blood tissues using publicly available databases.

2 | METHODS

2.1 | NF1 patients and CALM number in CALM1 and CALM2 cohorts

CALM1 and CALM2 studies were approved by the Institutional Review Boards (IRB) and the NCI Special Studies IRB. All participants provided written informed consent prior to joining the studies. The NCI Special Studies IRB also approved the overall GWAS. The study was carried out according to the principles of the Declaration of Helsinki.

Quantitative trait café-au-lait macule count data were collected from two cohorts, CALM1 (CALM number counted by DRS) and CALM2 (CALM number counted by AB), as described previously (Pemov et al., 2014). Most of the patients in the CALM1 cohort were adults, while most patients in the CALM2 cohort were children (Table 1). All patients in the CALM2 cohort underwent whole-body MRI and had at least one plexiform neurofibroma. In both cohorts, all CALMs ≥ 5 mm in longest diameter were counted.

TABLE 1   Descriptive information of European–American neurofibromatosis type 1 patients studied

	Genome-wide association study			RPS6KA2 Targeted sequencing		
	CALM1 (N = 112)	CALM2 (N = 59)	*p-value for t-test*[a]	CALM1[b] (N = 99)	CALM2[b] (N = 58)	*p-value for t-test*[a]
Age, years (Mean $\pm$ SD)	37.5 ± 14.0	14.5 ± 6.5	$< 2.2 \times 10^{-16}$	37.7 ± 14.4	14.5 ± 6.5	$<2.2 \times 10^{-16}$
Gender (male/female)	46/66	33/26	0.07	43/56	32/26	0.16
Body surface, m^2 (Mean $\pm$ SD)	1.8 ± 0.2	1.4 ± 0.4	1.3×10^{-11}	1.8 ± 0.3	1.3 ± 0.4	6.6×10^{-12}
CALM number (Mean $\pm$ SD)	20.9 ± 12.1	19.8 ± 10.9	0.56	21.4 ± 12.2	19.8 ± 11.0	0.41
Family size (number of families)	1, 2, 3, 4 (57, 16, 5, 2)	1, 2 (55, 2)		1, 2, 3, 4 (53, 13, 4, 2)	1, 2 (54, 2)	

Abbreviations: CALM, café-au-lait macule; RPS6KA2, ribosomal protein S6 kinase A2 gene; SD, standard deviation.

[a]p-values for two-sided, unequal variance t-test.

[b]information for the samples used for re-sequencing of intronic region of RPS6KA2 (NM_021135.6).

A Wood's lamp, which makes CALMs more visible, was used for CALM counting in CALM1, but not in CALM2. The log-transformed CALM number, $\log_{10}$ (CALM + 10), was analyzed adjusting for age, sex and body surface area $(m^2) = \sqrt{(\text{weight(kg)} \cdot \text{height(cm)})}/60$ defined as in Mosteller, 1987. Differences in age, sex, body surface, and CALM number were compared by two-sided t-test in R assuming unequal variance between cohorts.

2.2 | Genotypes for CALM1 samples for GWAS

Genotyping for the CALM1 samples was performed on Illumina Infinium HumanOmni-1Quad bead-chips. SNPs from 22 autosomes were included if the SNP GenTrain score was >0.3 and minor allele frequency (MAF) was >0 (non-monomorphic). Samples with genotyping rate $\geq$99% and calling rate $\geq$99% were retained. Cryptic relatedness and population stratification were checked with multidimensional scaling (MDS) analysis estimating pairwise identity by descent (IBD) for all pairs of founders using independent SNPs in PLINK (Purcell et al., 2007). SNPs with MAF $\geq$ 0.2 in founders were pruned for independence, first with pairwise linkage disequilibrium (LD) correlation measure $r^2 < 0.2$ within 50 SNPs, then again within 200 SNPs. Lastly, SNPs with MAF < 0.1 in European–Americans were excluded. The quality control protocol yielded 608,561 SNPs in 112 European–Americans with non-missing CALM number. The CALM1 cohort included 57 unrelated individuals, 16 families with 2 members each, 5 families with 3 members each, and 2 families with 4 members each.

2.3 | Genotypes for CALM2 samples for GWAS

Genotyping for the replication cohort CALM2 samples was performed on Illumina Infinium Omni2.5 bead-chips. Quality control measures were performed as described above although the retention rates for genotyping and calling rates were both lowered to $\geq$98% due to the smaller sample size. Samples with genotyping rate <95% from remaining European–Americans and SNPs with MAF < 0.2 were excluded—yielding 623,562 SNPs in 59 European–Americans with non-missing CALM number, including two sibling pairs. To ensure compatibility between the genotyping data between the two cohorts, LitfOver (Hinrichs et al., 2006) was used to convert genome positions in the CALM1 cohort (NCBI build 36, UCSC hg18) to corresponding positions in the CALM2 cohort (NCBI build 37 (UCSC hg19).

2.4 | Targeted sequencing of candidate loci

Targeted sequencing of candidate loci was carried out in 171 European–American samples using the AmpliSeq/IonTorrent platform, as per the manufacturer's instructions. A genotype was considered missing if its genotype quality score (GQ) was <10, its read depth (DP) score was <10 or the ratio of the genotype quality score (GQ) divided by genotype read depth (DP), GQ/DP was <0.5. Monomorphic SNPs and SNPs with a calling rate <90% and samples with a genotyping rate <90% were excluded leaving 176 SNPs in 157 samples. Quality control for targeted gene sequencing data was done using SNP & Variation Suite v8.3.4 (http://goldenhelix.com/products/SNP_Variation/index.html) and PLINK (Purcell et al., 2007).

2.5 | Tests of association

In each cohort, tests of association between each SNP and log-transformed CALM number, adjusted for age, sex, and body surface area, were performed with EMMAX (Kang et al., 2010) with an IBS kinship matrix considering possible spurious association due to genetic relatedness. Simple linear regression (SLR) ignoring genetic relatedness was also performed; these results are not shown because the correlations of the p-values for EMMAX and SLR for the CALM1 and CALM2 cohorts were 0.95 and 0.99, respectively. In the targeted gene sequencing samples, tests of association were performed by SLR between each SNP in *RPS6KA2* (OMIM 601685; GenBank NM_021135.6) and the log-transformed CALM number, adjusted for age, sex, and body surface area, for the CALM1 and CALM2 cohorts, both separately and combined. An additional covariate, cohort (CALM1, CALM2), as well as age, sex, and body surface area were adjusted in the mega analysis for combined-targeted gene sequencing data. Candidate loci were defined as those with suggestive genome-wide significance ($p < 1 \times 10^{-5}$) in the GWAS of either cohort or in the meta-analysis.

2.6 | Meta-analysis

Meta-analyses between the CALM1 and CALM2 cohort studies were performed with METAL (Willer, Li, & Abecasis, 2010) with weights proportional to the square-root of the sample size for each study (Stouffer, Suchman, Devinney, Star, & Williams, 1949), with genomic control correction of input statistics enabled. Heterogeneity analyses including I^2 statistics, the percentage of variation between cohorts due to heterogeneity rather than chance, were also performed in meta-analyses.

2.7 | ComPaSS-GWAS analysis

The second approach for corroboration of SNP associations used ComPaSS-GWAS (Sabourin et al., 2018), a method based on complementary pairs stability selection (Shah & Samworth, 2013) with a regression based GWAS analysis. ComPaSS-GWAS approximates replication by randomly splitting a sample in half (into a pseudo-discovery and pseudo-replication set) multiple times and looking for corroboration of results between each random split. ComPaSS-GWAS returns a corroboration score between 0 and 1, indicating the proportion of random splits where the SNPs were corroborated for the specified within-split significance parameter. The primary focus of ComPaSS-GWAS is for situations where appropriate independent replication data are either not available or when the replication cohort is too

small to sufficiently bolster the primary cohort's findings in a meta-analysis. ComPaSS-GWAS may also be useful when heterogeneity between the cohorts affects the ability of a meta-analysis to produce confident results. A high ComPaSS-GWAS score in a single cohort could also indicate that the lack of replication may be a result of the differences in cohorts rather than the individual cohort analysis result being a false positive. In this study, ComPaSS-GWAS scores were obtained using a within-split significance parameter of 10^{-3} based on 100 random splits. Simulations have shown (Sabourin et al., 2018) that a ComPaSS-GWAS score of 0.6 is comparable to a traditional genome-wide association significance level ($p < 5 \times 10^{-8}$). ComPaSS-GWAS score greater than 0.4 and less than 0.6 are considered to be suggestively significant findings.

2.8 | LD structures

Haploview 4.2 (Barrett, Fry, Maller, & Daly, 2005) was used to generate LD plots across the genome based on D' and r^2 values to evaluate LD structures among significant SNPs variants in Table 2 for each cohort.

2.9 | Functional annotation and quantitative-trait loci analysis of SNPs

Custom tracks on the UCSC Genome browser (http://genome. ucsc.edu) were used to screen NIH Roadmap Epigenomics and ENCODE experimental data containing the SNP regions of interest for evidence for regulatory relevance in normal skin cells (melanocytes, fibroblasts and keratinocytes). The online tools HaploReg (http://www.broadinstitute.org/ mammals/haploreg/haploreg.php) and RegulomeDB (http:// regulome.stanford.edu) were used as complementary analyses and to confirm the location of each SNP in relation to protein-coding and/or non-coding RNA genes. While many regulatory regions can act in a tissue-specific manner, it is unknown whether variants in such regions also have tissue-specific effects, and complex phenotypes are often caused by dysfunction of multiple tissues or cell type. This suggests the potential for the involvement of skin and other tissues/cells such as blood (immunity) and fibroblasts (regulatory role in pigmentation) in CALM phenotype. To evaluate if a SNP could influence gene expression *in-cis* or act as an expression quantitative trait loci (eQTL) in blood and skin tissues, all coding genes were identified that mapped within ±1 Mb of the index SNPs (associated SNPs of interest). eQTL results for all CALM-associated SNPs and collated genes based on 749 normal skin tissues (414 sun-exposed and 335 non-sun-exposed), 369 normal blood samples and 300 transformed fibroblast cell lines were then extracted from GTEx (V7,

TABLE 2 SNPs with suggestively genome-wide significant associations ($p < 1.0 \times 10^{-5}$) by EMMAX with log-transformed CALM number adjusting for age, sex and body surface area in CALM1 or CALM2 cohorts

SNP ID	Chr	Position (hg19)	Nearest neighbor gene	Type	MAF[a]	CALM1 EMMAX beta	CALM1 EMMAX p-value	CALM1 ComPaSS score	CALM2 EMMAX beta	CALM2 EMMAX p-value	CALM2 ComPaSS score	Meta p-value	Heterogeneity I^2 (%)	Heterogeneity p-value
rs4513317	2	101280697	*PDCL3, NPAS2*	Intergenic	0.357	0.018	3.43×10^{-1}	0	0.134	**6.61×10^{-6}**	*0.50*	6.80×10^{-4}	89.6	1.94×10^{-3}
rs7615501	3	53175557	*RFT1, PRKCD*	Intergenic	0.429	−0.030	1.27×10^{-1}	0	0.124	**7.53×10^{-7}**	*0.57*	9.55×10^{-2}	95.8	**9.88×10^{-7}**
rs4856490		84147884	*LINC02008, LINC00971*	Intergenic	0.143	0.116	**1.13×10^{-6}**	*0.54*	ND	ND	ND	ND	ND	ND
rs6872422	5	7422768	*ADCY2*	Intronic	0.370	−0.026	1.51×10^{-1}	0	0.107	**3.10×10^{-6}**	*0.43*	1.16×10^{-1}	95.3	**3.97×10^{-6}**
rs11742602		7440607			0.415	ND	ND	ND	0.125	**1.19×10^{-7}**	*0.75*	ND	ND	ND
rs1025291		7445289			0.305	−0.033	6.91×10^{-2}	0	0.124	**4.27×10^{-7}**	*0.67*	1.35×10^{-1}	96.2	**2.55×10^{-7}**
rs4710070	6	166995232	*RPS6KA2*	Intronic	0.292	0.096	**1.08×10^{-6}**	*0.57*	ND	ND	ND	ND	ND	ND
rs3799603		167001286			0.312	0.100	**7.40×10^{-8}**	*0.72*	−0.006	8.40×10^{-1}	0	2.27×10^{-5}	91.0	8.87×10^{-4}
rs3778385		167003381			0.299	0.096	**6.15×10^{-7}**	*0.63*	−0.0004	9.89×10^{-1}	0	5.63×10^{-5}	88.4	3.28×10^{-3}
rs12190451		167005259			0.312	0.100	**7.40×10^{-8}**	*0.72*	ND	ND	ND	ND	ND	ND
rs3799585		167016947			0.390	−0.086	**3.64×10^{-6}**	0.40	−0.034	2.48×10^{-1}	0	**9.31×10^{-6}**	68.0	7.70×10^{-2}
rs10946177		167018721			0.305	0.099	**4.34×10^{-7}**	*0.68*	0.027	3.71×10^{-1}	0	**3.94×10^{-6}**	80.2	2.48×10^{-2}
rs10215358	7	131363167	*PODXL, LOC101928782*	Intergenic	0.448	0.085	**5.84×10^{-6}**	*0.46*	−0.029	2.62×10^{-1}	0	2.62×10^{-3}	92.2	3.58×10^{-4}
rs12765442	10	25905421	*GPR158, LINC00836*	Intergenic	0.110	−0.124	**6.09×10^{-6}**	*0.51*	ND	ND	ND	ND	ND	ND
rs868692		25905810				ND	ND	ND	ND	ND	ND	ND	ND	ND
rs868691		25905876				ND	ND	ND	ND	ND	ND	ND	ND	ND
rs1951646	14	38146748	*TTC6*	Intronic	0.442	0.073	4.37×10^{-5}	0.28	0.061	4.67×10^{-2}	0.01	**7.67×10^{-6}**	0	4.28×10^{-1}
rs34047645	15	84611367	*ADAMTSL3*	Exonic	0.112	−0.116	**3.91×10^{-6}**	*0.53*	ND	ND	ND	ND	ND	ND
rs8093709	18	44134749	*LOXHD1*	Intronic	0.297	ND	ND	ND	0.144	**6.49×10^{-6}**	0.32	ND	ND	ND
rs732108		44135027			0.297	ND	ND	ND	ND	ND	ND	ND	ND	ND

Note: ND, no SNP data analyzed; p-values for EMMAX and Meta-analysis <1 × 10^{-5} and ComPaSS-GWAS score >0.4 are shown in bold and italic fonts, respectively. GenBank IDs for genes listed in the table, from top to bottom: NM_024065.5; NM_002518.4; NM_052859.4;NR_147146.1; NR_033860.1; NM_020546.3; NM_021135.6; NM_001018111.3; NR_110836.1; NM_020752.3; NR_108067.1; NM_001310135.2; NM_207517.3; NM_144612.6. If multiple transcripts exist for a gene, only an ID for transcript variant 1 is shown.

aMinor allele frequency in CALM1 samples or CALM2 samples if not in CALM1.

dbGaP Accession phs000424.v7.p2), (12 lead index SNPs and 1 surrogate variant in perfect LD). In addition, the influence of a SNP on metabolite levels in blood was evaluated by extraction of association results for each SNP and metabolites in blood from the Metabolomics GWAS (http://metabolomics.helmholtz-muenchen.de/gwas/).

3 | RESULTS

3.1 | GWAS for CALM1 and CALM2 cohorts

Descriptive information for the CALM1 and CALM2 cohorts and *p*-values comparing difference between cohorts by *t*-test analyses are presented in Table 1. The cohorts were significantly different with respect to age ($p = 2.2 \times 10^{-16}$) and body surface area ($p = 6.6 \times 10^{-12}$). Figure 1 presents the -$\log_{10}(p$-value) of tests of association using EMMAX on 608,561 SNPs from 112 samples in the CALM1 cohort (top panel) and 623,562 SNPs from 59 samples from the CALM2 cohort (middle panel). The bottom panel presents

-$\log_{10}(p$-value) from the meta-analysis of the 322,455 SNPs that were shared in common between the two cohorts. Table 2 presents SNPs that were suggestively genome-wide significant ($p < 1.0 \times 10^{-5}$) using EMMAX for either the CALM1 or CALM2 GWAS or the meta-analysis. *p*-values $< 1 \times 10^{-5}$ are shown in bold and italics.

In the CALM1 cohort GWAS, two SNPs in LD ($r^2 = 1$, D' = 1) on chromosome 6, rs12190451 and rs3799603, in the intronic region of *RPS6KA2*, had a borderline genome-wide significant association ($p = 7.4 \times 10^{-8}$) with CALM number (Figure S1a,b). Four additional significant SNP associations in the CALM1 cohort in this intronic region were significant at the suggestive genome-wide level ($p < 1.0 \times 10^{-5}$). Five of the six SNPs in *RPS6KA2* appeared to be in LD, but rs3799585 was not in high LD ($r^2 < 0.28$, D'=1) with the other SNPs (Figure S1a,b, Block 2). Figure 2 presents a regional visualization of EMMAX results for *RPS6KA2* in CALM1 samples (LocusZoom [Pruim et al., 2010]). In addition to the six intronic SNPs in *RPS6KA2*, other SNPs that were suggestively significant at a genome-wide association level included: an intergenic SNP on chromosome 3 (rs4856490, near *LINC00971*), an intergenic SNP on chromosome 7
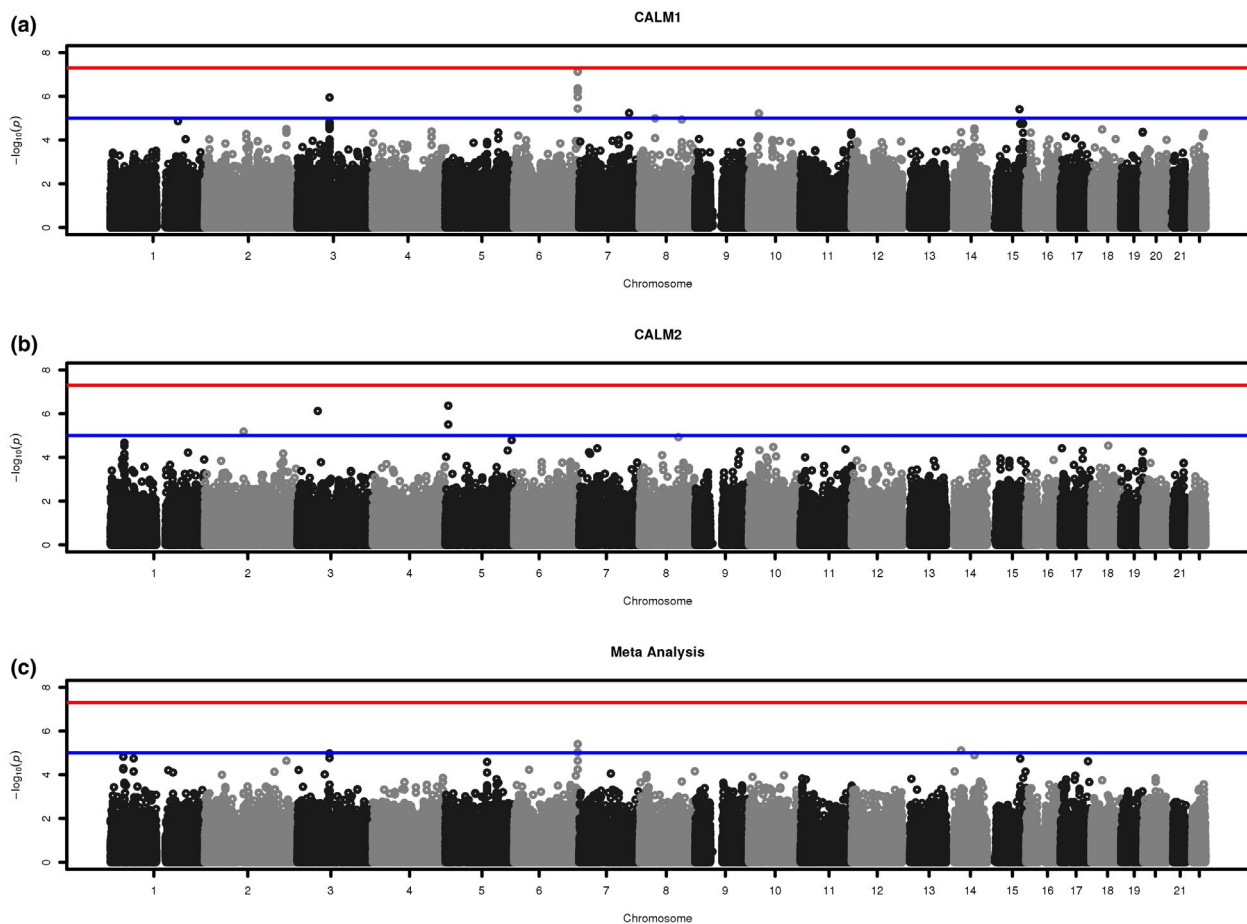

FIGURE 1 Manhattan plots of *p*-values obtained from the genome-wide EMMAX analyses. (a) CALM1 cohort; (b) CALM2 cohort; (c) Meta-analysis of CALM1 and CALM2. The red horizontal line corresponds to the genome-wide significance level of $p < 5 \times 10^{-8}$ and the blue line corresponds to suggestive genome-wide significance level of $p < 1.0 \times 10^{-5}$
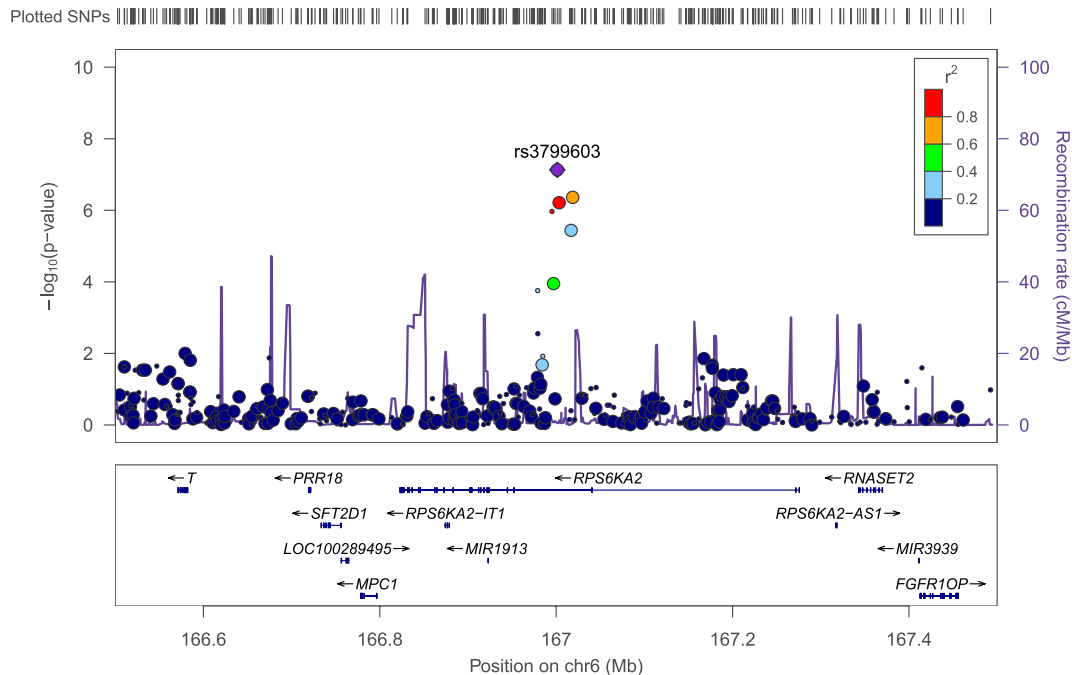
FIGURE 2 Regional visualization of EMMAX results for *RPS6KA2 locus* in CALM1 samples. The region in chromosome 6, including *RPS6KA2* and flanking genes is shown (genomic coordinates are in hg19). Individual SNPs depicted with circles. Negative log10 *p*-values and recombination rate are shown on the left and right vertical axes, respectively. Color code for r^2 values is shown in the inset on the right

(rs10215358) residing between *PODXL* and *LOC101928782*, three intergenic SNPs in LD ($r^2 = 1$, D'=1) between *GPR158* and *LINC00836* on chromosome 10 (rs12765442, rs868692, and rs868691) and an exonic SNP in *ADAMTSL3* on chromosome 15 (rs34047645).

Similarly, SNPs in the CALM2 cohort SNPs that were suggestively significant at the genome-wide level ($p < 1.0 \times 10^{-5}$) included one intergenic SNP on chromosome 2 (rs4513317 between *PDCL3* and *NPAS2*); one intergenic SNPs on chromosome 3 (rs7615501 between *RFT1* and *PRKCD*); three intronic SNPs in *ADCY2* on chromosome 5 (rs6872422, rs11742602, rs1025291); and two SNPs in LD ($r^2 = 1$, D'=1) in an intronic region of *LOXHD1* on chromosome 18 (rs8093709, rs732108). The SNPs in *ADCY2* and *LOXHD1* were in high LD within each gene (Figure S1c,d, Blocks 1 ($r^2 > 0.7$, D'=1) and 3 ($r^2 = 1$, D'=1)).

3.2 | Meta-analysis for CALM1 and CALM2

None of the SNPs that were significant at the suggestive level in the CALM1 GWAS were similarly significantly in the CALM2 GWAS. And conversely, none of the SNPs that were significant at the suggestive threshold in the CALM2 GWAS were similarly significantly in the CALM1 GWAS. Three SNPs were suggestively significant in the meta-analysis: rs3799585 and rs10946177 in the *RPS6KA2* gene on chromosome 6 and rs1951646 in the *TTC6* gene on chromosome 14.

Of the ten SNPs in common between the two cohorts, eight SNPs had an I^2 statistic, the proportion of the variance between the cohort due to heterogeneity rather than chance, greater than 80%, and three had an I^2 statistic greater than 95% ($p < 1 \times 10^{-5}$). And, in six of these eight SNPs, the regression coefficients (betas) were opposite in the direction between the two cohorts, although the differences were quite small.

3.3 | Cohort heterogeneity

In addition to the significant differences between the cohorts for age and body surface area (Table 1), there were striking differences between the cohorts in terms of the distribution of CALM number and age (Figure 3). CALM number decreased as age increased in the CALM1 cohort, but CALM number increased with age up to about 18 years of age in the CALM2 cohort. Linear regression modeling for CALM number versus age for two groups was performed (>18 years old versus ≤ 18 years old). The model *p*-value for the group of 124 NF1 patients older than 18 years of age (right side of the vertical dashed line) was 1.04×10^{-8} and the *p*-value for the group of 47 NF1 patients less than 18 years of age (left side of the vertical dashed line) was 0.04, confirming the opposite relationship between CALM number versus age by 18 years of age. Taking into consideration that most of the patients in the CALM1 cohort were older than 18 years and most of the patients in the CALM2 cohort were younger than 18 years
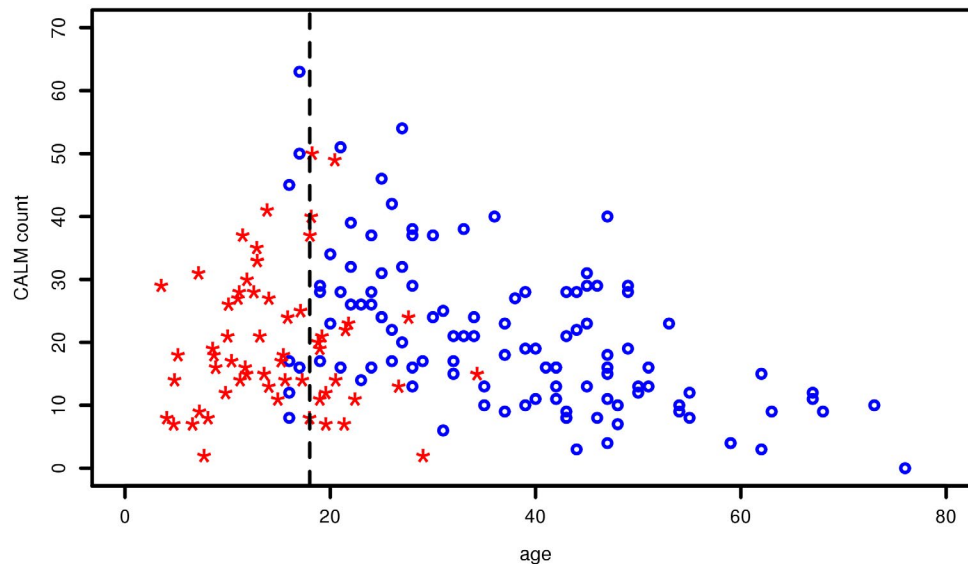
FIGURE 3 Comparison of café-au-lait macule number distribution by patients' age in CALM1 and CALM2 cohorts. Samples in CALM1 are depicted with open circles and samples in CALM2 are depicted with asterisks. Vertical dashed line is drawn at age of 18 years. Age on the X axis is shown in years

(Table 1), the opposite age distribution for CALM number could contribute to the opposing SNP effects observed in each independent GWAS. These findings suggest that the CALM2 GWAS was not an appropriate replication cohort for the CALM1 GWAS. To address this issue ComPaSS-GWAS was used on each cohort separately.

3.4 | ComPaSS-GWAS for CALM1 and CALM2

ComPaSS-GWAS corroboration scores using EMMAX were obtained for all SNPs, presented in Table 2. In the CALM1 cohort, the two most significant SNPs identified by EMMAX in *RPS6KA2* intron on chromosome 6 (rs3799603 and rs12190451) also had the highest ComPaSS-GWAS corroboration score (0.72). These SNPs were in LD ($r^2 = 1$; D'=1). Similarly, the two most significant SNPs identified by EMMAX in *ADCY2* on chromosome 5 in the CALM2 cohort (rs11742602 and rs1025291) also had the highest ComPaSS-GWAS corroboration scores of 0.75 and 0.67. The ComPaSS-GWAS analyses confirmed the findings by EMMAX within each cohort of CALM1 and CALM2 and suggest that these findings may be replicated if larger more homogeneous samples can be obtained.

3.5 | Targeted sequencing of *RPS6KA2*

Because of the marginally/borderline significant associations in the intronic region of *RPS6KA2*in the CALM1 cohort GWAS, ~24 Kb (chr6:166,995,059–167,018,915) of

intron 1 of *RPS6KA2* was sequenced (transcript variant 1, NM_021135.6) in order to investigate other possible variants. There were 15 *RPS6KA2* SNPs with suggestive genome-wide significance ($p < 1.0 \times 10^{-5}$) using simple linear regression (SLR) in the CALM1 cohort. However, no SNP was significant in the CALM2 cohort, or in meta- or mega-analyses of combined cohorts (Table S1).

3.6 | Functional annotation of CALM-associated SNPs

Using Roadmap and ENCODE experimental data from normal penile foreskin cells (including melanocytes, fibroblasts, and keratinocytes) and peripheral blood mononuclear cells, each of the CALM-associated SNPs at *p*-value < 1.0 x 10^{-5} presented in Table 2 were annotated and mapped to potential DNA regulatory regions (e.g., transcriptional start site [TSS], enhancers, and DNaseI/open chromatin sites; Table S2). *RPS6KA2* rs12190451, rs3778385, and rs4710070 mapped to either an active enhancer or region flanking an active enhancer in penile skin fibroblasts and/or melanocytes. Variation at rs4710070 also altered a CpG dinucleotide. *ADCY2* rs11742602 and *RPS6KA2* rs1025291were located in heterochromatin regions in fibroblasts, keratinocytes and melanocytes. The SNP rs7615501 on chromosome 3 was shown to map to an intergenic promoter region in fibroblasts. SNPs rs4513317 (chromosome 2), rs4856490 (chromosome 3), rs10215358 (chromosome 7), and rs12765442 (chromosome 10) were located in heterochromatin regions in fibroblasts, keratinocytes, and melanocytes. The SNP rs34047645 in exon 18 of *ADAMTSL3* and rs3799603 in intron 1 of *RPS6KA2*
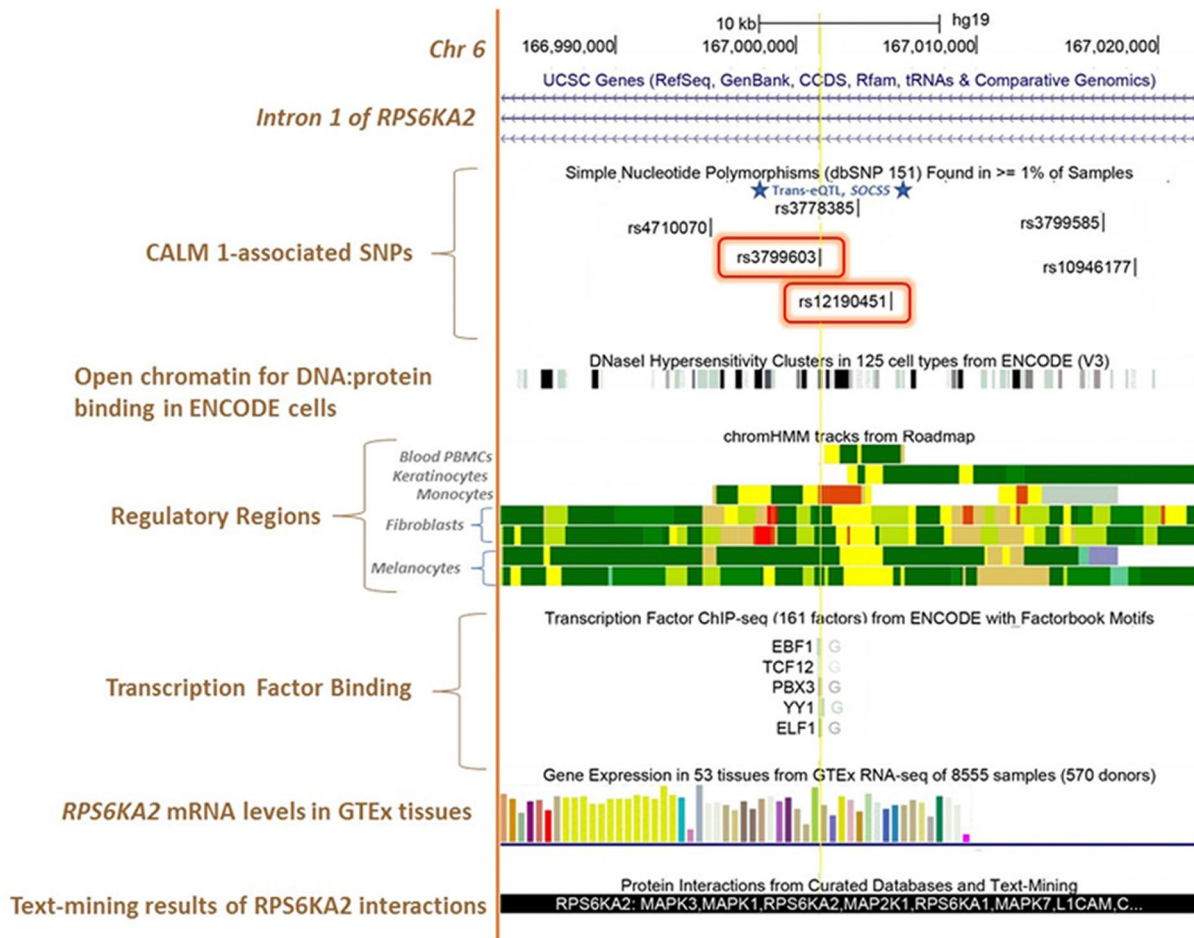
FIGURE 4 Genome browser image of intron 1 of *RPS6KA2* (NM_021135) on human assembly hg19 based on NIH Epigenomics Roadmap data and ENCODE data. *CALM1-associated SNPs in intron 1 of RPS6KA2:* ($p < 1.0 \times 10^{-5}$ threshold used) driven SNPs rs3799603 and rs12190451 that drive the significant association signal (include the *p*-value threshold cutoff here) observed in the intron are enclosed in red boxes. *Open Chromatin for DNA:protein binding in ENCODE Cells*: DNase I hypersensitivity clusters (solid black colored bars are presented) in 125 cell types from ENCODE. *Regulatory Regions*: chromatin state segmentation using a multivariate Hidden Markov Model (CHMM) and core histone marks are provided for blood cells (PBMCs), skin keratinocytes, monocytes, fibroblasts and melanocytes. Colored bars represent specific regulator regions: crimson, flanking promoter region; hot red, active promoter; lime green, enhancer-like region; dark grey, repressed polycomb region and light grey, weak repressed polycomb region. We show here that rs3778385 and rs3799603 (a trans-eQTL for *SOCS5* in monocytes, PMID: 24604202) locate to monocyte-specific promoter regions in open chromatin, but only rs3799603 (tracked by yellow line) can interact with transcription factors (e.g., ELF1 and TCF12). Also, rs12190451 locates to an enhancer-region in melanocytes. For clarity, all SNPs contained in this region according to dbSNP are not shown. *Transcription Factor Binding*: Chromatin immunoprecipitation sequencing (ChIP-seq; 161 factors) from ENCODE with FactorBook motifs highlighted by a green bar. *RPS6KA2 mRNA Levels in GTEx Tissues*: RNAseq expression in 53 tissues/8,555 sample. *Text-mining Results of RPS6KA2 Interactions*: Proteins interacting with *RPS6KA2* encoded protein, RSK3, are shown. Track sources and acknowledgements for the University of California Santa Cruz (UCSC) genome, ENCODE, the NIH Roadmap databases and extracted tracks (http://genome.ucsc.edu/goldenPath/credits.html#human_credits)

(Figure 4) did not locate to regulatory regions in skin or peripheral blood cells. SNP-containing DNA regions shown to interact with specific transcription factor proteins or alter protein binding motifs are also indicated in Table S2.

3.7 | QTL analysis of CALM-associated SNPs

The effect of each of the CALM-associated SNPs in Table 2 were further evaluated for effects on gene expression *in-cis*

using publicly available data from the National Institutes of Health Genotype-Tissue Expression GTEx data (V7) from normal whole blood, and skin (sun-exposed [lower leg] and not sun exposed [suprapubic]) tissues as well as transformed fibroblast cell lines (Table S3). Neither of the two genome-wide significantly-associated SNPs (rs3799603 and rs12190451) in CALM1were identified as potential eQTLs for *RPS6KA2* or other coding genes *in-cis*. Likewise, none of the variants identified by targeted gene sequencing of intron 1 of *RPS6KA2* in Table S1were eQTLs for *RPS6KA2* or other

genes *in-cis* (data not shown). rs11742602 and rs1025291 which showed the strongest association with CALM number in CALM2 were not eQTLs for *ADCY2* or other genes *in-cis* in the blood or skin tissues examined. In contrast, the C allele (Gly713Arg) of rs34047645 in exon 18 of *ADAMTSL3,* which was associated with CALM number in CALM1 results in significantly lower expression of *GOLGA6L4* (beta = −0.41, $p = 4.2 \times 10^{-5}$) in sun-exposed skin and higher expression of *CSPG4P11* (beta = 0.29, $p = 7.1 \times 10^{-5}$) in not sun-exposed skin. Also, the G allele of the intergenic SNP rs7615501on chromosome 3 which correlated with CALM number in CALM2 resulted in significantly higher expression of *RFT1* (beta 0.15, $p = 2.4 \times 10^{-5}$) in transformed fibroblasts and sun-exposed skin, and significantly lower expression of *GLT8G1* and *PHF7* in sun-exposed skin (Table S3). After correction for multiple comparisons, none of the remaining SNPs were eQTLs for any coding genes *in-cis* in the GTEx tissues examined (Table S3).

Using published association results (beta and *p*-values) from the genome-wide association studies (GWAS) of the human blood metabolome (http://metabolomics.helmholtz-muenchen.de/gwas/; Shin et al., 2014; Suhre et al., 2011), we further evaluated if any CALM-associated SNPs in Table 2 could act as metabolite QTLs (mQTLs) in blood (Table S4). After correcting for multiple comparisons (alpha of 0.05/20 tests = *p*-value ≤ 2.5×10^{-3}), *RPS6KA2* SNPs rs4710070 and rs10946177 ($r^2 = 0.83$) were significant mQTLs for glutamine (beta = 0.01, *p*-value = 9.8×10^{-4}) and pro-lyl-4-hydroxyproline (beta = 0.02, *p*-value = 6.1×10^{-4}), respectively. In addition, rs12765442, rs868692, and rs868691 on chromosome 10, which were in LD ($r^2 = 1$, D'=1) were mQTLs for uridine (beta =−0.022, *p*-value = 1.9×10^{-4}) and hypoxanthine (beta = −0.032, *p*-value = 3.7×10^{-4}) in blood.

4 | DISCUSSION

To identify common variants that influence the number of café-au-lait macules in individuals with NF1, a GWAS was performed in two cohorts with CALM number being treated as a quantitative trait. In the CALM1 cohort, multiple, significant SNPs were identified in the intronic regions of *RPS6KA2* (but not in CALM2) and in the CALM2 cohort SNPs were identified in *ADCY2* (OMIM103071; but not in CALM1). Meta-analysis of the two cohorts identified no genome-wide level significant SNPs, although three SNPs (two in *RPS6KA2* (rs3799585 and rs10946177)) and one SNP in *TTC6* (rs1951646) were significant at a suggestive genome-wide association level (*p*-value < 1×10^{-5}). Post-genotyping analysis of the cohorts showed significant differences in age and body surface area, differences between the distribution of CALM number and age, and a large number of the variants shared in common between the two cohorts suggested that differences may likely be due to heterogeneity rather than type I error.

To address this heterogeneity issue, Complimentary Pairs Stability Selection—Genome-Wide Association Study or ComPaSS-GWAS (Sabourin et al., 2018), a method that approximates replication by randomly splitting the data into pseudo-discovery and pseudo-replication sets, was used to identify SNPs in each cohort separately. ComPaSS-GWAS identified four SNPs with EMMAX in the CALM1 cohort (in *RPS6KA2*) and two SNPs in the CALM2 cohort (in *ADCY2*) with ComPaSS-GWAS corroboration score >0.6. Based on previous simulation studies, this suggests that these findings could be replicated in larger and/or more homogeneous samples.

The only SNPs that had a significant ComPaSS-GWAS corroboration score (>0.6), were located in the introns of *RPS6KA2* and *ADCY2*. These two genes belong to RAS-MAPK and cyclic AMP pathways, respectively. These two pathways interact in the cell via common members and converge on the family of transcription factors that control gene expression by binding to cAMP responsive element (CREBs), found in the promotor regions of a number of genes. Both MAPK and cAMP pathways are known to play important roles in proliferation, differentiation, apoptosis and DNA repair among other cellular functions (Anjum & Blenis, 2008; Houslay, 2006; Sassone-Corsi, 2012).

RPS6KA2 (ribosomal protein S6 kinase A2, legacy name *RSK3*) encodes a 90 kDa protein that belongs to a family of closely related (75%–80% amino-acid identity) and evolutionary conserved serine-threonine kinases that are phosphorylated and activated by RAS-MAPK pathway kinases, encoded by *MAPK3* (OMIM601795) and *MAPK1* (OMIM 176948; Anjum & Blenis, 2008). *RPS6KA2* has been identified as a modifier of epidermal growth factor receptor activity in pancreatic cancer (Milosevic et al., 2013). Although no known human phenotypes are associated with *RPS6KA2*, *RPS6KA3* (OMIM 300075) underlies Coffin-Lowry syndrome (CLS; Trivier et al., 1996), a rare X-linked disorder of cognitive delay and skeletal malformations. However, no pigmentary abnormalities are reported in CLS.

The cAMP pathway has been shown to be involved in NF1 pathogenesis. *ADCY2* (adenylate cyclase 2) is a member of a family of genes that encode enzymes catalyzing synthesis of cyclic adenosine 3',5'-monophosphate (cAMP) from adenosine triphosphate (ATP; Sassone-Corsi, 2012). In mammalian brain, NF1/RAS may affect cellular cAMP through protein kinase C zeta and cell-type specific G-protein coupled receptors (Anastasaki & Gutmann, 2014). Additionally, a study of 243 NF1 patients that focused on association of 2,761 unique SNPs in 22 key regulators of intracellular cAMP levels with optic pathway glioma (OPG), identified an adenylate cyclase

gene, *ADCY8* (OMIM 103070), to be significantly associated with OPG risk (Warrington et al., 2015).

Functional annotation of CALM-associated SNPs in this study revealed that the majority of *RPS6KA2* SNPs including the significantly associated rs12190451 in CALM1 mapped to potential enhancers or enhancer-like regions and/or resided in strong open chromatin regions in primary skin cells. The *ADCY2* SNPs located to heterochromatin and/or polycomb-repressed regions in fibroblasts and keratinocytes, but not melanocytes. However, none of the *RPS6KA2* or *ADCY2* associated SNP (and with ComPaSS-GWAS scores >0.6) were identified as eQTLs for the respective genes or any genes *in-cis* in sun-exposed, non-exposed skin tissues, whole blood, and transformed cultured fibroblasts from GTEx (Table S3). The possibility remains, however, that these SNPs may influence mRNA expression of these genes in melanocytes as both RSK2 and ADCY2 proteins are expressed in melanocytes (https://www.proteinatlas.org/).

RPS6KA2 rs3799585 which was associated with CALM number in CALM1 (p-value = 4.34×10^{-7}), but not CALM2, has been reported to be a trans-eQTL for *SOCS5* (OMIM607094) on chromosome 2 in monocytes (p-value = 3.15×10^{-8}; Fairfax et al., 2014) suggesting that the other haplotype SNPs may also influence monocyte levels of *SOCS5* mRNA. We show here using RoadMap and Encode data that rs3799585 together with the significant signal rs3799603 (p-value = 7.4×10^{-8}) locate to a promoter region in monocytes, but not in skin cells, and unlike rs3799585, rs3799603 can influence the binding of several transcription factors including YY1 (Choudhury & Ramsey, 2016). SOCS5 is a member of the suppressor of cytokine signaling (SOCS) family of proteins and plays an important role in several immune response pathways including IL4 and JAK-STAT signaling. IL4 (OMIM 147780) can directly inhibit melanogenesis in normal healthy melanocytes via STAT3 (OMIM 102582) and STAT6 (OMIM 601512) phosphorylation and downregulate both transcription and translation of melanogenesis-associated genes, such as microphthalmia-associated transcription factor and dopachrome tautomerase (Fairfax et al., 2014). Thus, the involvement of IL4-induced JAK2-STAT6 signaling in melanocytes could be considered a potential mechanism for influencing pigmentation in skin.

We also observed that *RPS6KA2* rs4710070 and rs10946177 ($r^2 = 0.83$) can influence the level of glutamine and prolyl-4-hydroxyproline metabolites, respectively, in blood (Table S4). Both SNPs were associated with increased CALM number in CALM1 only. Proliferating cells display an intense appetite for glutamine, reflecting its versatility as a nutrient and mediator of intermediary metabolism (DeBerardinis & Cheng, 2010). A role for glutamine in extracellular signal-regulated protein kinase (ERK) signaling/pathways has also been identified in intestinal epithelial cells (DeBerardinis & Cheng, 2010). Prolyl-hydroxyproline

(Pro-Hyp) has also been shown to stimulate the growth of skinfibroblasts in rats and consequently increases fibroblasts migrating from the skin (Shigemura et al., 2009).

The lack of replication in this study may likely be due to heterogeneity between discovery (CALM1) and replication (CALM2) cohorts. Several factors may be responsible. The data from this study show an increase in the CALM number with age up to ~18 years and a decrease in CALM number after 18 years (Figure 3). CALM number is known to increase in early childhood (Boyd et al., 2010; Shah, 2010) and then decrease with age (Duong et al., 2011; Huson, Harper, & Compston, 1988). Clinically different features of the patients of the two cohorts due to ascertainment bias of recruitment could explain other parts of the heterogeneity. For the CALM1 cohort, all NF1 patients were eligible, regardless of specific phenotypic features. For the CALM2 cohort, only NF1 patients with plexiform neurofibromas were recruited. Since plexiform neurofibromas can be associated with overlying hyper-pigmentation (Shah, 2010), the high prevalence (essentially 100%) of these tumors in CALM2 that may have inflated the CALM number. Furthermore, plexiform neurofibromas tend to grow during childhood and especially puberty (Dombi et al., 2007; Nguyen et al., 2012); the associated dyspigmentation may also follow this trend. This may partially explain the increase in CALM number we observed in patients <18 years of age; the mini-peak in CALM number around 11–12 years of age is particularly notable (Figure 3).

Difference in phenotyping of NF1 traits is another possible cause of heterogeneity between cohorts. The Wood's lamp method (UV light in a dark room) method was used to find CALM for CALM1; this is helpful to find fainter CALM in older people. However, counting of CALM by Wood's lamp was not used for CALM2. Accurate and robust quantitative phenotyping of clinical traits in NF1 is technically challenging, highly time- and labor-consuming and remains one of the major obstacles in identification of genetic modifiers in NF1.

Another possible interpretation of the lack of replication could be that multiple genetic modifiers are involved in NF1, each with a limited effect on the phenotype, but few of them could be identified in a modestly sized GWAS and none in a small one. However, as ComPaSS-GWAS results suggest, present findings within each cohort might be biologically plausible given larger or more homogeneous cohorts.

In summary, we performed the first GWAS for CALM number in neurofibromatosis type 1. We used café-au-lait macule number as a tumor-like and easily quantifiable clinical phenotype to identify genome-wide variation that is associated with this trait. We identified SNPs in intronic region of *RPS6KA2*, a member of RAS-MAPK signaling pathway, that were associated with CALM number at a suggestively significant genome-wide level. *RPS6KA2* is biologically plausible as a genetic modifier of NF1 and warrants further investigation and functional studies.

CONFLICT OF INTERESTS

DRS performs contract clinical telegenetics services for Genome Medical, Inc, in accordance with relevant NCI ethics policies.

AUTHOR CONTRIBUTIONS

HS and PLH analyzed data, interpreted results, wrote and prepared manuscript for publication. AP managed clinical samples, wrote and prepared manuscript for publication, and submitted manuscript for publication. JAS analyzed data. AMB collected clinical samples and data. SB oversaw high-throughput genotyping and sequencing of DNA. KT and WL analyzed high-throughput genotyping and sequencing data. FNLCR (collectively) processed, SNP-array genotyped and sequenced DNA samples. BCW oversaw clinical portion of the study. DRS originated the study, collected and analyzed clinical data, contributed to experimental design, and wrote the manuscript. AFW contributed to experimental design, oversaw data analyses, and wrote the manuscript.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Heejong Sung https://orcid.org/0000-0002-2329-326X
Alexander Pemov https://orcid.org/0000-0003-3893-3685
Douglas R. Stewart https://orcid.org/0000-0001-8193-1488

REFERENCES

Anastasaki, C., & Gutmann, D. H. (2014). Neuronal NF1/RAS regulation of cyclic AMP requires atypical PKC activation. *Human Molecular Genetics*, *23*(25), 6712–6721. https://doi.org/10.1093/hmg/ddu389

Anjum, R., & Blenis, J. (2008). The RSK family of kinases: Emerging roles in cellular signalling. *Nature Reviews Molecular Cell Biology*, *9*(10), 747–758. https://doi.org/10.1038/nrm2509

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265. https://doi.org/10.1093/bioinformatics/bth457

Boyd, K. P., Gao, L., Feng, R., Beasley, M., Messiaen, L., Korf, B. R., & Theos, A. (2010). Phenotypic variability among cafe-au-lait macules in neurofibromatosis type 1. *Journal of the American Academy of Dermatology*, *63*(3), 440–447. https://doi.org/10.1016/j.jaad.2009.09.042

Brems, H., Beert, E., de Ravel, T., & Legius, E. (2009). Mechanisms in the pathogenesis of malignant tumours in neurofibromatosis type 1. *The Lancet Oncology*, *10*(5), 508–515. https://doi.org/10.1016/s1470-2045(09)70033-6

Brems, H., Park, C., Maertens, O., Pemov, A., Messiaen, L., Upadhyaya, M., … Stewart, D. R. (2009). Glomus tumors in neurofibromatosis type 1: Genetic, functional, and clinical evidence of a novel association. *Cancer Research*, *69*(18), 7393–7401. https://doi.org/10.1158/0008-5472.Can-09-1752

Carey, J. C., & Viskochil, D. H. (1999). Neurofibromatosis type 1: A model condition for the study of the molecular basis of variable expressivity in human disorders. *American Journal of Medical Genetics*, *89*(1), 7–13. https://doi.org/10.1002/(SICI)1096-8628(19990326)89:1<7:AID-AJMG4>3.0.CO;2-#

Choudhury, M., & Ramsey, S. A. (2016). Identifying cell type-specific transcription factors by integrating ChIP-seq and eQTL data-application to monocyte gene regulation. *Gene Regulation and Systems Biology*, *10*, 105–110. https://doi.org/10.4137/grsb.S40768

DeBerardinis, R. J., & Cheng, T. (2010). Q's next: The diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*, *29*(3), 313–324. https://doi.org/10.1038/onc.2009.358

Dombi, E., Solomon, J., Gillespie, A. J., Fox, E., Balis, F. M., Patronas, N., … Widemann, B. C. (2007). NF1 plexiform neurofibroma growth rate by volumetric MRI: Relationship to age and body weight. *Neurology*, *68*(9), 643–647. https://doi.org/10.1212/01.wnl.0000250332.89420.e6

Duong, T. A., Bastuji-Garin, S., Valeyrie-Allanore, L., Sbidian, E., Ferkal, S., & Wolkenstein, P. (2011). Evolving pattern with age of cutaneous signs in neurofibromatosis type 1: A cross-sectional study of 728 patients. *Dermatology*, *222*(3), 269–273. https://doi.org/10.1159/000327379

Easton, D. F., Ponder, M. A., Huson, S. M., & Ponder, B. A. (1993). An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): Evidence for modifying genes. *American Journal of Human Genetics*, *53*(2), 305–313.

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., … Knight, J. C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, *343*(6175), 1246949. https://doi.org/10.1126/science.1246949

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., … Kent, W. J. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*, *34*(90001), D590–D598. https://doi.org/10.1093/nar/gkj144

Houslay, M. D. (2006). A RSK(y) relationship with promiscuous PKA. *Science Signaling*, *2006*(349), pe32. https://doi.org/10.1126/stke.3492006pe32

Huson, S. M. (2008). The neurofibromatoses: classification, clinical features and genetic counseling. In D. Kaufmann (Ed.), *Monographs in human genetics*, (Vol. *16*, pp. 1–20). Basel: Karger.

Huson, S. M., Harper, P. S., & Compston, D. A. (1988). Von Recklinghausen neurofibromatosis. A clinical and population study in south-east Wales. *Brain*, *111*(6), 1355–1381. https://doi.org/10.1093/brain/111.6.1355

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–354. https://doi.org/10.1038/ng.548

Maertens, O., De Schepper, S., Vandesompele, J. O., Brems, H., Heyns, I., Janssens, S., … Messiaen, L. (2007). Molecular dissection of isolated disease features in mosaic neurofibromatosis type 1.

American Journal of Human Genetics, *81*(2), 243–251. https://doi.org/10.1086/519562

Milosevic, N., Kühnemuth, B., Mühlberg, L., Ripka, S., Griesmann, H., Lölkes, C., … Michl, P. (2013). Synthetic lethality screen identifies RPS6KA2 as modifier of epidermal growth factor receptor activity in pancreatic cancer. *Neoplasia*, *15*(12), 1354–1362. https://doi.org/10.1593/neo.131660

Mosteller, R. D. (1987). Simplified calculation of body-surface area. *New England Journal of Medicine*, *317*(17), 1098. https://doi.org/10.1056/nejm198710223171717

Nguyen, R., Dombi, E., Widemann, B. C., Solomon, J., Fuensterer, C., Kluwe, L., … Mautner, V.-F. (2012). Growth dynamics of plexiform neurofibromas: A retrospective cohort study of 201 patients with neurofibromatosis 1. *Orphanet Journal of Rare Diseases*, *7*, 75. https://doi.org/10.1186/1750-1172-7-75

Pemov, A., Sung, H., Hyland, P. L., Sloan, J. L., Ruppert, S. L., Baldwin, A. M., … Stewart, D. R. (2014). Genetic modifiers of neurofibromatosis type 1-associated cafe-au-lait macule count identified using multi-platform analysis. *PLoS Genetics*, *10*(10), e1004575. https://doi.org/10.1371/journal.pgen.1004575

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., … Willer, C. J. (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, *26*(18), 2336–2337. https://doi.org/10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Sabbagh, A., Pasmant, E., Laurendeau, I., Parfait, B., Barbarot, S., Guillot, B., … Wolkenstein, P. (2009). Unravelling the genetic basis of variable clinical expression in neurofibromatosis 1. *Human Molecular Genetics*, *18*(15), 2768–2778. https://doi.org/10.1093/hmg/ddp212

Sabourin, J. A., Cropp, C. D., Sung, H., Brody, L. C., Bailey-Wilson, J. E., & Wilson, A. F. (2018). ComPaSS-GWAS: A method to reduce type I error in genome-wide association studies when replication data are not available. *Genetic Epidemiology*, *43*(1), 102–111. https://doi.org/10.1002/gepi.22168

Sassone-Corsi, P. (2012). The cyclic AMP pathway. *Cold Spring Harbor Perspectives in Biology*, *4*(12), a011148. https://doi.org/10.1101/cshperspect.a011148

Shah, K. N. (2010). The diagnostic and clinical significance of cafe-au-lait macules. *Pediatric Clinics of North America*, *57*(5), 1131–1153. https://doi.org/10.1016/j.pcl.2010.07.002

Shah, R. D., & Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society*, *75*(1), 55–80. https://doi.org/10.1111/j.1467-9868.2011.01034.x

Sharif, S., Moran, A., Huson, S. M., Iddenden, R., Shenton, A., Howard, E., & Evans, D. G. (2007). Women with neurofibromatosis 1 are at a moderately increased risk of developing breast cancer and should be considered for early screening. *Journal of Medical Genetics*, *44*(8), 481–484. https://doi.org/10.1136/jmg.2007.049346

Shigemura, Y., Iwai, K., Morimatsu, F., Iwamoto, T., Mori, T., Oda, C., … Sato, K. (2009). Effect of Prolyl-hydroxyproline (Pro-Hyp), a food-derived collagen peptide in human blood, on growth of fibroblasts from mouse skin. *Journal of Agriculture and Food Chemistry*, *57*(2), 444–449. https://doi.org/10.1021/jf802785h

Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., … Soranzo, N. (2014). An atlas of genetic influences on human blood metabolites. *Nature Genetics*, *46*(6), 543–550. https://doi.org/10.1038/ng.2982

Stewart, D. R., Corless, C. L., Rubin, B. P., Heinrich, M. C., Messiaen, L. M., Kessler, L. J., … Brooks, D. G. (2007). Mitotic recombination as evidence of alternative pathogenesis of gastrointestinal stromal tumours in neurofibromatosis type 1. *Journal of Medical Genetics*, *44*(1), e61. https://doi.org/10.1136/jmg.2006.043075

Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams, R. M. Jr (1949). *The American soldier: adjustment during army life. (Studies in social psychology in World War II)* (Vol. *1*). Oxford, England: Princeton University Press.

Suhre, K., Shin, S.-Y., Petersen, A.-K., Mohney, R. P., Meredith, D., Wägele, B., … Gieger, C. (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, *477*(7362), 54–60. https://doi.org/10.1038/nature10354

Szudek, J., Joe, H., & Friedman, J. M. (2002). Analysis of intrafamilial phenotypic variation in neurofibromatosis 1 (NF1). *Genetic Epidemiology*, *23*(2), 150–164. https://doi.org/10.1002/gepi.1129

Trivier, E., De Cesare, D., Jacquot, S., Pannetier, S., Zackai, E., Young, I., … Hanauer, A. (1996). Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature*, *384*(6609), 567–570. https://doi.org/10.1038/384567a0

Uusitalo, E., Rantanen, M., Kallionpää, R. A., Pöyhönen, M., Leppävirta, J., Ylä-Outinen, H., … Peltonen, J. (2016). Distinctive cancer associations in patients with neurofibromatosis type 1. *Journal of Clinical Oncology*, *34*(17), 1978–1986. https://doi.org/10.1200/jco.2015.65.3576

Warrington, N. M., Sun, T., Luo, J., McKinstry, R. C., Parkin, P. C., Ganzhorn, S., … Rubin, J. B. (2015). The cyclic AMP pathway is a sex-specific modifier of glioma risk in type I neurofibromatosis patients. *Cancer Research*, *75*(1), 16–21. https://doi.org/10.1158/0008-5472.Can-14-1891

Welti, S., D'Angelo, I., & Scheffzek, K. (2008). Structure and function of neurofibromin. In D. Kaufmann (Ed.), *Neurofibromatoses. monographs in human genetics* (Vol. *16*, pp. 113–128). Basel: Karger.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.