Arthritis & Rheumatology
Vol. 76, No. 1, January 2024, pp 141–145
DOI 10.1002/art.42673

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Genome-Wide Association Study Identifies the First Germline Genetic Variant Associated With Erdheim-Chester Disease

Javier Martínez-López,[1] Ana Márquez,[1] Francesco Pegoraro,[2] Lourdes Ortiz-Fernández,[1] Marialbert Acosta-Herrera,[3] Martin Kerick,[1] Elena Gelain,[4] Eli L. Diamond,[5] Benjamin H. Durham,[5] Omar Abdel-Wahab,[5] Ronald S. Go,[6] Matthew J. Koster,[6] Lorenzo Dagna,[7] Corrado Campochiaro,[7] Matthew Collin,[8] Paul Milne,[8] Juvianee I. Estrada-Veras,[9] Kevin O'Brien,[9] Matthias Papo,[10] Fleur Cohen-Aubar,[11] Zahir Amoura,[10] Julien Haroche,[10] Javier Martín,[1] and Augusto Vaglio[2]

Objective. Erdheim-Chester disease (ECD) is rare histiocytosis with a wide range of clinical manifestations. Somatic mutations are key to the pathogenesis of the disease; however, the relationship between germline genetic variants and ECD has not been examined so far. The present study aims to explore the inherited genetic component of ECD by performing the first genome-wide association study.

Methods. After quality controls, a cohort of 255 patients with ECD and 7,471 healthy donors was included in this study. Afterward, a logistic regression followed by in silico functional annotation was performed.

Results. A signal at the 18q12.3 genomic region was identified as a new susceptibility locus for ECD ($P = 2.75 \times 10^{-11}$; Odds Ratio = 2.09). This association was annotated to the *SETBP1* gene, which is involved in clonal haematopoiesis. Functional annotation of this region and of the identified suggestive signals revealed additional genes that could be potentially involved in the pathogenesis of the disease.

Conclusion. Overall, this work demonstrates that germline genetic variants can impact on the development of ECD and suggests new pathways with a potential pathogenic role.

INTRODUCTION

Erdheim-Chester disease (ECD) is a rare histiocytosis with approximately 1,500 reported cases worldwide.[1] It mostly affects middle-aged individuals[2] and has a heterogeneous phenotype, which ranges from indolent to multisystemic and aggressive forms; disease-related morbidity and mortality are still considerable, especially in patients with central nervous system (CNS) involvement.[2]

This research is part of the doctoral degree awarded to Mr. Martínez-López, within the Biomedicine program from the University of Granada.

Supported by the project "Genome-wide and epigenome-wide association study in patients with Erdheim-Chester Disease," funded by Histio UK, and supported in part by the Intramural Research Program of the National Human Genome Research Institute and Redes de Investigación Cooperativa Orientadas a Resultados en Salud (RD21/0002/0039). Dr. Ortiz-Fernández's work was supported by a Juan de la Cierva Incorporación fellowship (IJC2019-040746-I) funded by MCIN/AEI/10.13039/501100011033. Dr. Accosta-Herrera's work was supported by a Miguel Servet fellowship (CP21/00132) from Instituto de Salud Carlos III.

[1]Javier Martínez-López, MSc, Ana Márquez, PhD, Lourdes Ortiz-Fernández, PhD, Martin Kerick, PhD, Javier Martín, MD, PhD: Institute of Parasitology and Biomedicine López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain; [2]Francesco Pegoraro, MD, PhD, Augusto Vaglio, MD, PhD: University of Florence and Meyer Children's Hospital Instituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Florence, Italy; [3]Marialbert Acosta-Herrera, PhD: Institute of Parasitology and Biomedicine López-Neyra, Consejo Superior de Investigaciones Científicas and Hospital Clínico San Cecilio, Ibs. Granada Instituto de Investigación Biosanitaria, Granda, Spain; [4]Elena Gelain, PhD: Meyer Children's Hospital Instituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Florence, Italy; [5]Eli L. Diamond, MD, Benjamin H. Durham, MD, Omar Abdel-Wahab, MD: Memorial Sloan Kettering Cancer Center, New York, New York; [6]Ronald S. Go, MD, Matthew J. Koster, MD: Mayo Clinic, Rochester, Minnesota; [7]Lorenzo Dagna, MD, Corrado Campochiaro, MD: Instituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Raffaele Scientific Institute and Vita-Salute San Raffaele University, Milano, Italy; [8]Matthew Collin, MD, PhD, Paul Milne, MD, PhD: Newcastle University, Newcastle upon Tyne, UK; [9]Juvianee I. Estrada-Veras, MD, Kevin O'Brien, MS, CRNP: National Human Genome Research Institute, NIH, Bethesda, Maryland; [10]Matthias Papo, MD, PhD, Zahir Amoura, MD, Julien Haroche, MD, PhD: Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Service de Médecine Interne and Centre National de Référence Maladies Systémiques Rares et Histiocytoses, Hôpital Pitié-Salpêtrière, Centre d'immunologie et des maladies infectieuses INSERM UMRS1135, Paris, France; [11]Fleur Cohen-Aubar, MD: Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Service de Médecine Interne, Paris, France.

Drs. Martínez-López and Márquez are co-first authors and contributed equally to this work. Drs. Martín and Vaglio are co-senior authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/10.1002/art.42673).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/art.42673.

Address correspondence via email to Javier Martín MD, PhD, at javiermartin@ipb.csic.es.

Submitted for publication March 31, 2023; accepted in revised form August 7, 2023.

KEY POINTS

1. The 18q12.3 genomic region is a new susceptibility locus for Erdheim-Chester disease; this association is annotated to the *SETBP1* gene.
2. Inherited variants are relevant in the development of Erdheim-Chester disease.

ECD features both neoplastic and inflammatory components[3]; it is thought to be driven by the clonal expansion of mutated histiocytes that migrate to tissues, acquire a senescent phenotype, and secrete proinflammatory and profibrotic cytokines and chemokines, which promote tissue damage.[3] The somatic mutations that induce histiocyte proliferation involve the MAPK and PI3K-AKT pathways.[4,5] $BRAF^{V600E}$ is the most common mutation; it occurs in approximately 55% to 60% of patients[3,4] and has a strong pathogenic relevance underlined by the therapeutic efficacy of its inhibitor vemurafenib.

The molecular basis of ECD, however, has not been fully elucidated. In this regard, understanding the genetic variation underlying disease susceptibility using genome-wide approaches could provide relevant insights into disease pathogenesis. Indeed, analyses of inherited variants in ECD-related disorders, such as Langerhans cells histiocytosis, has yielded interesting results.[6] This work aimed to explore the potential role of germline genetic variation in ECD susceptibility by performing the first genome-wide association study (GWAS) in this disease.

MATERIALS AND METHODS

Blood samples from 315 patients with ECD and 7,911 healthy donors of European ancestry from four different countries (Italy, France, United Kingdom, and United States) were analyzed (Supplementary Table 1). This study was approved by the CSIC Ethic Committee and the Ethic Committee of Research of the Granada Province, and written informed consent was obtained in accordance with the tenets of the Declaration of Helsinki.

Genotyping was performed using the arrays specified in Supplementary Table 1. PLINK 1.9 (https://www.cog-genomics.org/plink/) was used for the statistical analysis. Quality control (QC) filters were applied as follows: single nucleotide polymorphisms (SNPs) with call rates less than 0.95, minor allele frequencies less than 0.05, and those deviating from Hardy-Weinberg equilibrium ($P < 0.001$) were filtered out. Samples with call rates less than 0.95, first-degree relatives (Pi_Hat > 0.40), and duplicates (Pi_Hat > 0.99) were removed. QC-filtered GWAS data were imputed using the multiancestry TOPMed reference panel (https://imputation.biodatacatalyst.nhlbi.nih.gov/#!pages/home). Imputation results with rsq less than or equal to 0.9 were removed from further analyses. After imputation, 116,316 independent SNPs were used to perform a principal component (PC) analysis. First 10 PCs per individual were calculated, and those individuals showing greater than 4 SDs from the cluster centroid were excluded.

Genome-wide association analysis was performed using a logistic regression model, including 10 first PCs as covariates to avoid population bias. The genomic inflation factor (λ) was calculated to evaluate population stratification. Genome-wide significant threshold was established at P less than 5×10^{-8} and suggestive associations at P less than 1×10^{-5}. The presence of independent signals was assessed by conditional analysis, including the lead SNP as a covariate using GCTA64 and R base software.

The potential functional impact of associated variants was assessed in silico by exploring databases of chromatin interactions and expression quantitative trait loci (eQTL) using FUMA (https://fuma.ctglab.nl/). Moreover, GARFIELD was used to perform regulatory enrichment analysis to evaluate overlap between associated and suggestive variants and histone marks, associated with promoters, enhancers, and active genes, in a wide range of cell types and tissues (https://www.ebi.ac.uk/birney-srv/GARFIELD/). Finally, gene prioritization analysis of regions showing a suggestive level of association was performed by MAGMA via FUMA (https://ctg.cncr.nl/software/magma). Functional analysis parameters are detailed in the Supplementary Data.

Summary statistics of this study will be available under reasonable request.

RESULTS

Following QC and imputation analysis, 7,726 individuals (255 ECD and 7,471 healthy donors) and 5,178,225 genetic variants were included in the logistic regression model. No evidence of population stratification was observed (λ = 1.03).

After association analysis, a genomic association at the 18q12.3 region was identified (Figure 1 and Table 1). Within this region, 51 SNPs reached the genome-wide significance threshold. The most significant genetic variant was rs188050966 ($P = 2.75 \times 10^{-11}$). The minor allele (A) of this SNP was associated with a two-fold increased risk of ECD (odds ratio [OR] 2.09; 95% CI 1.68–2.6) (Table 1). Furthermore, analysis conditioned on the lead SNP revealed no secondary signals within the associated loci (Supplementary Figure 1), because no SNP reached the genome-wide significance threshold, nor the suggestive, after conditioning to the lead SNP of this region.

The lead SNP is located in an intergenic region with *SETBP1* being the closest gene. Functional annotation of rs188050966, and its proxies ($r^2 > 0.6$), prioritized seven coding genes, including *SETBP1*, as potential targets of this signal (Figure 2). Four genes (*HOXB2*, *GATD3*, *LAMB2*, *SH3PXD2A*) were prioritized based on the overlap of proxy SNPs with eQTLs in whole

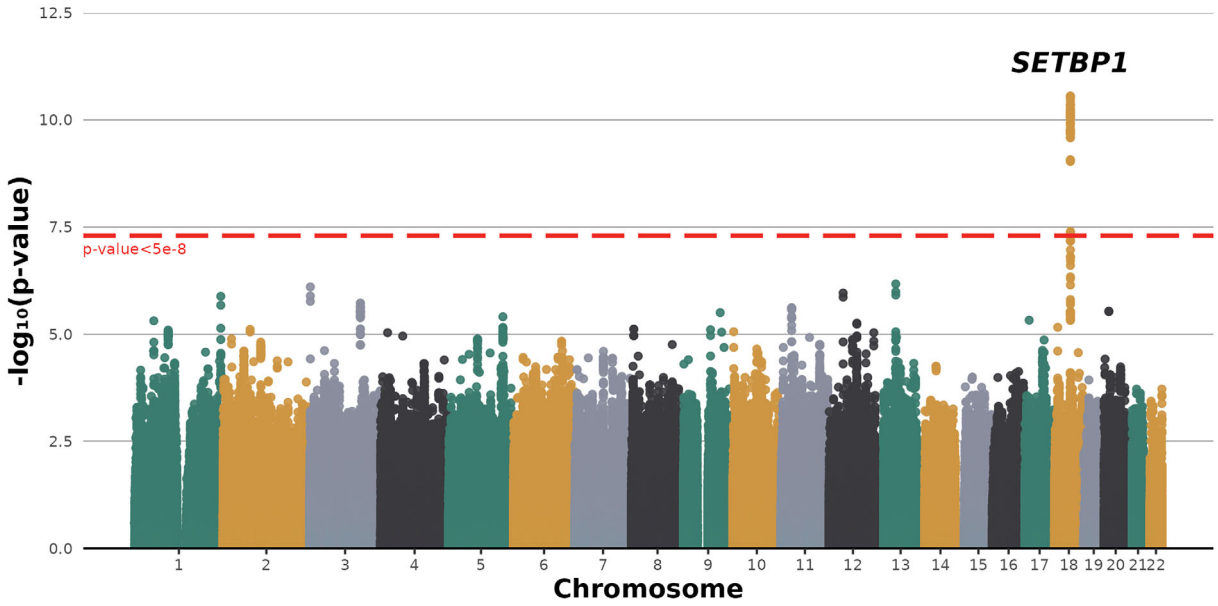

Figure 1. Manhattan plot representing the results of the genome-wide association study. *P* values calculated as the –log10 are plotted against the physical chromosomal position of each variant. The red line indicates the genome-wide level of significance ($P < 5 \times 10^{-8}$) and the blue line indicates the suggestive level of significance ($P < 1 \times 10^{-6}$).

blood (Supplementary Table 2). Specifically, the minor allele of rs16978075 (which correlated with the minor allele of the lead SNP, $r^2 = 0.78$) appeared to decrease gene expression levels of *LAMB2* and *SH3PXD2A*, whereas the minor allele of rs11082383 (correlated with the minor allele of the lead SNP, $r^2 = 0.83$) was associated with lower expression of *LAMB2*, *GATD3*, and *SH3PXD2A*, and increased expression of *HOXB2*. This last genetic variant also reached genome-wide significance in the association analysis ($P = 2.15 \times 10^{-10}$). Furthermore, three additional genes (*PIK3C3*, *SETBP1*, *SLC14A2*) showed evidence of physical interaction with the associated region (Figure 2).

Additionally, 11 regions showed a suggestive level of significance (Table 1). After gene prioritization analysis, 12 genes were suggested as the most probable causal genes at these loci (Table 1).

Moreover, the regulatory enrichment analysis, including both genome-wide and suggestive associated variants, revealed an enrichment of regulatory histone marks across 15 distinct cell types from eight tissue categories (Supplementary Figure 2). This enrichment was especially significant in blood, bone, and skin, which we found an overlap with different histone marks representing promoters (H3K9ac and H3K4me2), enhancers (H2BK20ac, and H3K27ac), and active genes (H2BK15ac and H3K79me1).

Table 1. Genetic variants reaching genome-wide or suggestive association with Erdheim-Chester disease*

Chr	Bp	Locus	SNP	EA	MAF		*P* value	OR (95% CI)	Func. refgene
					ECD	HD			
18	**44498986**	***SETBP1***	**rs188050966**	**A**	**0.23**	**0.13**	**2.75E-11**	**2.09 (1.68–2.60)**	**Intergenic**
13	53527973	*OLFM4*	rs2208989	CA	0.25	0.16	1.02E-06	1.71 (1.38–2.11)	Intergenic
1	242656412	*PLD5*[a]	rs10926753	C	0.47	0.42	1.31E-06	1.56 (1.30–1.87)	Intronic
3	141489115	*ATOX1*[a,b] */ G3BP1*[a]	rs73872319	G	0.09	0.05	1.89E-06	2.16 (1.57–2.96)	Intergenic
11	30484012	*MPPED2*[a]	rs11031101	A	0.20	0.12	2.43E-06	1.74 (1.38–2.18)	Intronic
20	12990837	*SPTLC3*[a]	rs360528	A	0.20	0.14	2.90E-06	1.72 (1.37–2.15)	Intronic
5	151147042	*CSMD1*	rs10062143	A	0.27	0.18	3.93E-06	1.61 (1.32–1.98)	Intronic
1	55972500	*Y_RNA*	rs356119	T	0.20	0.12	4.88E-06	1.69 (1.35–2.12)	Intergenic
8	6168673	*PCSK5*	rs57296216	C	0.10	0.05	7.66E-06	1.85 (1.49–2.29)	Intergenic
2	75779513	*EVA1A*[a]	rs12053280	T	0.42	0.32	7.75E-06	1.52 (1.26–1.82)	Intronic
1	96073713	*RP11-286B14.1*	rs543539	T	0.21	0.14	8.03E-06	1.67 (1.33–2.09)	ncRNA-intronic
12	125363662	*SCARB1*[a]	rs73229512	T	0.20	0.12	9.25E-06	1.69 (1.34–2.13)	Intergenic

* Genome-wide significant locus for ECD is highlighted in bold. 95% CI, 95% confidence interval; Bp, base pair; Chr, chromosome; EA, effect allele; ECD, Erdheim-Chester disease; HD, healthy donors; MAF, minor allele frequency in the study cohort; OR, odds ratio; SNP, single nucleotide polymorphism.
[a] Prioritized genes by MAGMA.
[b] Most significant gene after prioritization.

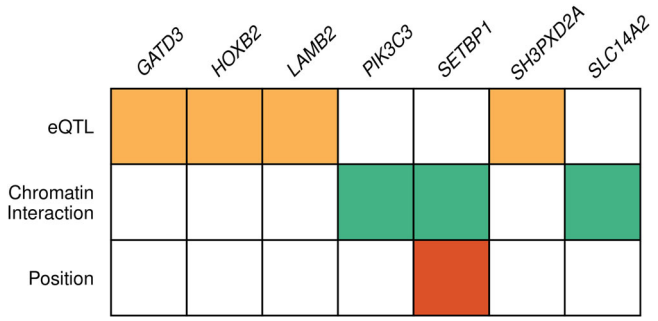

Figure 2. In silico annotation of the associated signal at the 18q12.3 region performed with FUMA. eQTL, expression quantitative trait loci. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.42673/abstract.

DISCUSSION

This study identified the first genomic region associated with susceptibility to ECD ($P = 2.75 \times 10^{-11}$). Compared with the ORs reported in complex diseases, the effect of the lead SNP rs188050966 (OR = 2.09) on ECD is quite strong. The closest gene to this region is *SETBP1*, which encodes a protein involved in proliferation of myeloid cells and, interestingly, *SETBP1* has been described to potentially influence the development of myeloid malignancies.[7] Even though no somatic mutations in *SETBP1* have been reported in ECD,[8] this gene has been associated with susceptibility to clonal haematopoiesis,[9] which is a major process in ECD pathogenesis.[8] However, given the limited sample size and the lack of specificity of this process in ECD, additional analyses are required to validate the association between this genomic locus and clonal hematopoiesis in ECD. Moreover, SNPs within this locus have been associated with changes in blood cell counts in previous studies.[10,11] Additionally, *SETBP1* has also been associated with the mosaic loss of Y chromosome, (mLOY)[9] a frequent phenotype in clonal hematopoiesis that could be intriguing to explore in ECD. Altogether, these findings suggest a potential role of *SETBP1* in the increased histiocyte proliferation occurring in ECD.

Regarding in silico annotation analysis, the most significant eQTL affected gene expression levels of *LAMB2* [$P = 3.98 \times 10^{-13}$] encoding a laminin involved in cell proliferation, migration, and differentiation. Although this protein is mostly expressed in other tissues,[12] its role in blood cells may be worth exploring. Additionally, the protein encoded by *PIK3C3*, which physically interacts with the associated SNPs, is regulated by the mammalian target of rapamycin (mTOR) complex,[13] which is involved in ECD pathogenesis.[5] Despite that *SETBP1* could be considered the best candidate target gene of this signal, our results reveal new genes potentially affecting the development of ECD that should be considered in follow-up studies to determine the accurate molecular mechanism through which this locus affects the disease.

Furthermore, 12 more genes were related to genomic regions with a suggestive level of significance. Among them, it is worth mentioning the role played by *SCARB1* in the metabolism of myeloid cells[14] and the relationship between the *ATOX1* copper transporter and the MAPK signaling pathway.[15]

Because of the broad spectrum of clinical manifestations and affected tissues in ECD, we hypothesized that germline genetic variants could have an impact on different cell types. Concerning this, we found an enrichment of the set of significant and suggestive SNPs overlapping histone marks not only in blood but also in other tissues affected by the disease, such as bone, skin, and CNS (Supplementary Figure 2), suggesting that molecular studies in nonblood tissues could provide a more accurate view of ECD pathogenesis.

Our study has constraints, namely the small sample size, which limits its statistical power. However, ECD is extremely rare, and it must be acknowledged that we analyzed approximately 20% of the global population of patients with reported ECD. Our analysis detected more than 10 suggestive associations that did not reach the genome-wide significance threshold. Their potential role in ECD remains uncertain and should be explored in studies with higher statistical power.

In conclusion, this study reports the first association between an inherited genetic variant and ECD, underscoring the relevance of germline variation in the development of this disease. Identification of novel pathways involved in ECD could have a substantial impact on the management of the disease. Indeed, molecular studies in ECD have achieved a fruitful translation into clinical practice, as targeted therapies, such as B-Raf proto-oncogene (BRAF) and Mitogen-Activated Protein Kinase (MEK) inhibitors, have greatly improved patient outcome.[2]

ACKNOWLEDGMENTS

We thank Sofia Vargas and Gema Robledo for their excellent technical assistance and all the patients and healthy donors for their essential collaboration.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Martín had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Martín, Vaglio.

Acquisition of data. Pegoraro, Gelain, Diamond, Durham, Abdel-Wahab, Go, Koster, Dagna, Campochiaro, Collin, Milne, Estrada-Veras, O'Brien, Papo, Cohen-Aubar, Amoura, Haroche.

Analysis and interpretation of data. Martínez-López, Márquez, Ortiz-Fernández, Acosta-Herrera, Kerick.

REFERENCES

1. Haroche J, Cohen Aubart F, Amoura Z. Erdheim-Chester disease. Blood 2020;135(16):1311–8.
2. Goyal G, Heaney ML, Collin M, et al. Erdheim-Chester disease: consensus recommendations for evaluation, diagnosis, and treatment in the molecular era. Blood 2020;135(22):1929–45.
3. Pegoraro F, Papo M, Maniscalco V, et al. Erdheim-Chester disease: a rapidly evolving disease model. Leukemia 2020;34(11):2840–57.
4. Emile JF, Diamond EL, Hélias-Rodzewicz Z, et al. Recurrent RAS and PIK3CA mutations in Erdheim-Chester disease. Blood 2014;124(19): 3016–9.
5. Gianfreda D, Nicastro M, Galetti M, et al. Sirolimus plus prednisone for Erdheim-Chester disease: an open-label trial. Blood 2015;126(10): 1163–71.
6. Peckham-Gregory EC, Chakraborty R, Scheurer ME, et al. A genome-wide association study of LCH identifies a variant in SMAD6 associated with susceptibility. Blood 2017;130(20):2229–32.
7. Makishima H, Yoshida K, Nguyen N, et al. Somatic SETBP1 mutations in myeloid malignancies. Nat Genet 2013;45(8):942–6.
8. Cohen Aubart F, Roos-Weil D, Armand M, et al. High frequency of clonal hematopoiesis in Erdheim-Chester disease. Blood 2021; 137(4):485–92.
9. Kessler MD, Damask A, O'Keeffe S, et al. Common and rare variant associations with clonal haematopoiesis phenotypes. Nature 2022; 612(7939):301–9.
10. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 2016;167(5):1415–29.e19.
11. Kanai M, Akiyama M, Takahashi A, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet 2018;50(3):390–400.
12. Matejas V, Hinkes B, Alkandari F, et al. Mutations in the human laminin beta2 (LAMB2) gene and the associated phenotypic spectrum. Hum Mutat 2010;31(9):992–1002.
13. Munson MJ, Allen GF, Toth R, et al. mTOR activates the VPS34-UVRAG complex to regulate autolysosomal tubulation and cell survival. EMBO J 2015;34(17):2272–90.
14. Tenesaca S, Vasquez M, Fernandez-Sendin M, et al. Scavenger receptor class B type I is required for 25-hydroxycholecalciferol cellular uptake and signaling in myeloid cells. Mol Nutr Food Res 2020; 64(15):e1901213.
15. Brady DC, Crowe MS, Turski ML, et al. Copper is required for oncogenic BRAF signalling and tumorigenesis. Nature 2014;509(7501): 492–6.