Endemic Burkitt Lymphoma in second-degree relatives in Northern Uganda: in-depth genome-wide analysis suggests clues about genetic susceptibility

Mateus H. Gouveia1, Isaac Otim2, Martin D. Ogwang3, Wang Mingyi4, Bin Zhu4, Nathan Cole4, Wen Luo4, Hicks Belynda4, Kristine Jones4, Kathrin Oehl-Huber5, Leona W. Ayers6, Stefania Pittaluga7, Ismail D. Legason8, Hadjah Nabalende2, Patrick Kerchan9, Tobias Kinyera2, Esther Kawira9, Glen Brubaker10, Arthur Levin9, Lutz Guertler11, Jung Kim12, Douglas Stewart12, Melissa Adde13, Ian Magrath13, Andrew W Bergen12, Steven J. Reynolds14, Meredith Yeager4, Kishor Bhatia12, Adebowale A. Adeyemo1, Ludmila Prokunina-Olsson12, Michael Dean12, Daniel Shriner1, Charles N. Rotimi1, Stephen Chanock12, Reiner Siebert5,*, Sam M. Mbulaiteye12,*

1Center for Research on Genomics & Global Health, NHGRI, National Institutes of Health, Bethesda, Maryland, United States of America 2EMBLEM Study, African Field Epidemiology Network, Kampala, Uganda 3EMBLEM Study, St. Mary’s Hospital, Lacor, Gulu, Uganda 4Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer 5Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, Germany 6Department of Pathology, The Ohio State University, Columbus, Ohio 7Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 8EMBLEM Study, Kuluva Hospital, Arua, Uganda 9EMBLEM Study, Shirati Health, Education, and Development Foundation, Shirati, Tanzania 10Inter-Church Medical Assistance Mission, Baltimore, Maryland, USA 11Max von Pettenkofer Institute, LMU University of München, Germany 12Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, United States of America 13International Network for Cancer Treatment, Brussels, Belgium 14Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, United States of America

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: Sam M. Mbulaiteye, mbulaits@mail.nih.gov.

*These authors also contributed equally to this work.

Competing Interests: The authors declare no competing interests.

Availability of data and materials: The datasets generated and/or analyzed during the current study are available through dbGAP (The EMBLEM data are available through accession: phs001705.v2.p1; the Shirati data are available through accession: phs002223.v1.p1; the Childhood Cancer Survivorship study are available through accession: phs002072.v1.p1; The International Cancer Genome Consortium (ICGC) data were extracted from WGS alignments are available from the European Genome-phenome archive (EGA) under the accession numbers: EGA-S00001002198 in accordance with approval from the ICGC guidelines (www.icgc.org) under DACO-1064755 (National Institutes of Health).
To the Editor:

Burkitt lymphoma (BL) is an aggressive B-cell lymphoma known to occur as endemic, sporadic, and immunodeficiency types. Endemic Burkitt lymphoma (eBL) has been linked to Plasmodium falciparum malaria and Epstein-Barr virus infections. Regardless of type, BL is characterized by hallmark somatic IG-MYC chromosomal translocations, which cause BL in conjunction with other somatic genetic or epigenetic abnormalities, including in TP53, ID3, and TCF3. However, the role of germline genetic susceptibility in BL has not been well studied, although it has been suspected based on reports of familial BL clusters.

We report detailed genome-wide analyses in Northern Uganda, including whole-exome sequencing (WES) of eBL tumor and germline DNA (Fig. S1), and genome-wide genotyping array (for Methods see Supplementary data). Based on > 4 million variants (Illumina 5M array) in 198 eBL cases enrolled in the epidemiology of Burkitt lymphoma in East African children and minors (EMBLEM) case-control study (2010–2016), we incidentally discovered two children (one boy and one girl aged 10–15 years) with eBL to be second-degree relatives, likely half-siblings (IBD proportion = 0.28, Fig. 1A).

Using the 198 representative children with confirmed eBL in Northern Uganda as a denominator, we estimate that about 1% of the eBL cases may be genetically related. The two children lived near each other in a high malaria transmission area (Fig. 1A). They presented with a short history of symptoms that started three months apart and were diagnosed with eBL based on histological criteria (Fig. 1B): consistent morphology on hematoxylin and eosin stains, positivity for B cell and germinal center markers, high proliferation index, EBV RNA and MYC translocation. Results of genome-wide copy number variation (CNV) analyses of the tumor DNA based on OncoScan SNP-array (Fig. 1C) and WES (Fig. S2) were consistent with reports that BL has a simple karyotype pattern. Specifically, the genomic imbalance mapping of the tumors revealed only a few alterations besides deletions at the IG loci, suggestive of clonal IG rearrangement (Table S1). In the first patient, these changes include apparent copy number neutral loss of heterozygosity (CNN-LOH) in 17p, spanning TP53, and a heterozygous loss in 18q21.32, spanning CCBE1. Alterations in the second patient were consistent with IG-rearrangement associated loss and a putative low level (sub-clonal) gain in 1q as highly recurrent in BL (Fig. 1C).

This discovery of BL in the two close relatives triggered a review of their medical records, which confirmed Stage C high-risk eBL in both patients. Both responded to treatment (INCTR 03–06 protocol) and were cured (Table 1). Both patients reported their paternal tribe as Langi, which belongs to the Western Nilotic ethno-linguistic group. Consistently, ADMIXTURE analysis showed that both children have more than 50% Nilotic ancestry (Fig. 1A).

We analyzed ~400 000 variants in germline WES and 282 mutations identified in tumor WES of the two children (Fig. S1). We focused on candidate germline variants defined as those with moderate-to-high deleterious Combined Annotation Dependent Depletion
(CADD) score 9 or those mapped to genes that are recurrently mutated in BL tumors (n=61) or in other cancers (n=180) (Table S2) 5,10.

We identified 106 404 identical-by-state (IBS) germline variants (~¼ of the total) in the two children, in agreement with their estimated second-degree of genetic relatedness. Of these variants, 254 were rare variants, with minor allele frequency (MAF) ≤0.01 in the ancestrally similar (Nilotic) reference population from Shirati, Tanzania (Table S3 and S4), and 784 were uncommon variants based on a less stringent MAF ≤0.05 (Table S5). Fifteen of the rare variants had a phred-scaled CADD scores > 10 (range 10.3 – 25.3) (Table S3). All the variants were validated by manually reviewing using The Integrative Genomics Viewer (IGV; Fig. S3A). None of the 15 variants were in a region previously suggested to be of regulatory importance in BL based on analysis of differential states of chromatin accessibility in BL-derived cell lines or non-neoplastic B-cells, or showed differential DNA methylation in BL 11 (Table S6). From the 241 candidate genes (Table S2), we identified an intronic SNP (rs772535596) in CHD8 (Table S3 and S4), which had a phred-scaled CADD score of ~10, consistent with “moderate evidence of pathogenicity” 12,13. CHD8 was recently identified as recurrently mutated in BL in the BL Genome Sequencing Project (BLGSP) 4. The rs772535596 SNP was not observed in the germline DNA of the unrelated Nilotic individuals from Shirati with BL (n=30) and without BL (n=80).

We also discovered 14 candidate variants in 13 genes in the germline DNA of both children (Table S3 and S4), based on a phred-scaled CADD score >10. Most of these variants were classified as variant of unknown significance (VUS), benign, or likely benign, based on InterVar. Most of the variants were rare (MAF< 1% in the gnomAD database). Of potential importance was an intronic indel rs374301928-ATT/- upstream of exon 12/20 (NM_001369568) in the TCF4 gene. This variant had a phred-scaled CADD score of 18.7 and “supporting evidence of pathogenicity” according to VarSome12,13, although it is currently classified as VUS by InterVar. This variant was observed in only one of the unrelated BL Nilotic patients from Shirati but not in any of the unrelated healthy Nilotic individuals from Shirati (p = 0.02, Fisher exact test). The rs374301928 variant was observed only nine (0.1%) times in 4 357 African genomes in the gnomAD database (Table S3), suggesting that this is a rare African-specific variant. The variant is found in a highly conserved genomic region, with predicted transcription factor binding sites adjacent to exon 12/20 (Fig. S4). Of interest, the rs374301928-TCF4 locus appears to have been subject to early negative selection among vertebrate species and archaic hominins that may be indicative of relevant regulatory function 14. Of pathological relevance to BL, TCF4 encodes the helix-loop-helix transcription factor 4 reported to interact with ID3, which is inactivated by recurrent mutations in up to two-thirds of BL 6,7,15. Somatic TCF4 deregulation has been implicated as an alternative mechanism of ID3-inactivating or TCF3-activating mutations in BL 5,15. Our somatic WES analysis showed that both patients lacked mutations in ID3 and TCF3. Since TCF4 has been implicated in the ID3/TCF3 pathway in BL 6,7,15, the observed germline/somatic pattern in our cases raises the question of whether germline TCF4 genetic variants could have an effect comparable to somatic involvement of ID3/TCF3. In view of their potential significance, both the CHD8 and TCF4 germline variants were verified by Sanger sequencing (Fig. S3B).
The somatic WES analysis for these two children identified 29 mutations in core genes, including *CCND3*, *MYC*, and *USP7* in one child and *BCL7A* and *DDX3X* in the other child, that have been reported to be recurrently mutated in other BL studies (Table S7)\(^4^\)\(^-\)\(^7\). Most (266) of the somatic mutations were unique to each child’s tumor (Tables S7 and S8). In addition to the mutations identified in the candidate BL genes, we identified 253 mutations in genes that have not been reported before in BL or other cancers; 70 of these mutations had phred-scaled CADD scores > 20 (Table S8).

While the discovery of genetic relatedness in these two eBL cases suggests a possible genetic predisposition to BL, environmental predisposition from *P. falciparum* malaria and EBV was considered. Both children did not carry common malaria-resistance genetic variants (e.g., the sickle cell trait, see Supplementary data)\(^3\). One child carried the HLA-B53 allele, previously reported to be associated with resistance to severe malaria in West Africa\(^16\). Both children were EBV-tumor positive and positive for EBV LMP-1 DNA Pattern A variant (Table 1), which has been associated with a 31-fold higher odds of eBL in EMBLEM \(^17\). However, the relatively advanced age of BL diagnosis in these children (>10 years) and recent efforts to suppress malaria in their district casts doubt on the hypothesis that these environmental pathogens are the sole triggers.

Our study is a discovery effort with several strengths, including epidemiologically well-characterized samples, extensive genomic data, and availability of tumor tissue. These strengths enabled us to robustly confirm diagnosis by histology and molecular analysis, and to conduct integrated multi-disciplinary analysis combining somatic and germline WES data. We confirmed genetic relatedness, ancestry and discovered two variants that warrant follow-up. However, the limitations include small sample size, which precluded consideration of formal statistics (including adjustment for multiple comparisons), and the lack of functional validation of the variants. Also, the paucity of genomic data from individuals in the eBL belt, i.e., Nilotic speakers \(^2\)\(^-\)\(^3\), is a limitation. Our study illustrates the feasibility and scalability of collaborative efforts applying genomic data analysis to identify familial aggregation of eBL in epidemiological or clinical cohorts and that such investigations can shed light on the genetic susceptibility to eBL.

In conclusion, we report the first pathologically confirmed eBL cases in Northern Uganda determined to be related based on their genetic data uncovered and analyzed in the course of an epidemiological study. We identified in both children potentially important germline DNA genetic variants in *TCF4* and *CHD8*. These discoveries, although preliminary, provide novel clues about genetic susceptibility to eBL development.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgements**

We thank the study subjects for their participation. We thank Dr. Joseph Fraumeni Jr, former Director of the Division of Cancer Epidemiology and Genetics (DCEG) at National Cancer Institute (NCI), for his support and scientific guidance while conducting this study. We thank Ms. Janet Lawler-Heavner at Westat Inc. (Rockville, MD, USA) and Mr. Erisa Sunday at the African Field Epidemiology Network (Kampala, Uganda) for managing the

*Leukemia. Author manuscript; available in PMC 2021 April 13.*
study. We thank Ms. Laurie Buck, Dr. Carol Giffen, and Mr. Greg Rydzak at Information Management Services Inc. (Calverton, MD, USA) for preparing data analysis files. We thank the members of the tumor genetic group at the Institute of Human Genetics for expert technical assistance and fruitful discussions and Josef Högel for support with CADD score evaluation. The study was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI) (Contracts HHSN261201100063C and HHSN261201100071), the Intramural Research Program, National Institute of Allergy and Infectious Diseases (SIR) and Center for Research on Genomics & Global Health, National Human Genome Research Institute, National Institutes of Health (NIH), Department of Health and Human Services, and, in part, by intramural funds of the Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, Germany. The infrastructural support of the KinderKrebsInitiative Buchholz/Holm-Seppensen to the group at the Institute of Human Genetics in Ulm is gratefully acknowledged. The authors acknowledge the research contributions of the Cancer Genomics Research Laboratory for their expertise, execution, and support of this research in the areas of project planning, wet laboratory processing of specimens, and bioinformatics analysis of generated data. from the NCI, NIH, under NCI Contract No. 75N91D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. The content of this manuscript is the sole responsibility of the authors.

References

1. Leoncini L, Campo E, Stein H, Harris NL, Jaffe ES, Kluin PM. Burkitt-like lymphoma with 11q aberration. WHO classification of tumours of haematopoietic and lymphoid tissues Revised 4th ed France, Lyon: IARC 2017; : 334.
2. Gouveia MH, Bergen AW, Borda V, Nunes K, Leal TP, Ogwang MD et al. Genetic signatures of gene flow and malaria-driven natural selection in sub-Saharan populations of the ‘endemic Burkitt Lymphoma belt’. PLoS Genet 2019; 15: e1008027. [PubMed: 30849090]
3. Legason ID, Pfeiffer RM, Udquim K-I, Bergen AW, Gouveia MH, Kirimunda S et al. Evaluating the Causal Link Between Malaria Infection and Endemic Burkitt Lymphoma in Northern Uganda: A Mendelian Randomization Study. EBiomedicine. 2017; 25: 58–65. [PubMed: 29033373]
4. Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. Blood 2019; 133: 1313–1324. [PubMed: 30617194]
5. López C, Kleinheinz K, Aukema SM, Rohde M, Bernhart SH, Hübßchmann D et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. Nat Commun 2019; 10: 1459. [PubMed: 30926794]
6. Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet 2012; 44: 1316–1320. [PubMed: 23143595]
7. Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 2012; 490: 116–120. [PubMed: 22885699]
8. Morrow RH, Pike MC, Smith PG, Ziegler JL, Kisuule A. Burkitt’s Lymphoma: A Time-space Cluster of Cases in Bwanba County of Uganda. Br Med J 1971; 2: 491–492. [PubMed: 4325452]
9. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019; 47: D886–D894. [PubMed: 30371827]
10. Waszak SM, Tiao G, Zhu B, Rausch T, Muyas F, Rodriguez-Martin B et al. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. bioRxiv. 2017.https://edoc.mdc-berlin.de/17576/ (accessed 14 Jun2019).
11. Kretzmer H, Bernhart SH, Wang W, Haake A, Weniger MA, Bergmann AK et al. DNA methylene analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. Nat Genet 2015; 47: 1316–1325. [PubMed: 26437030]
12. Kopanos C, Tsiolkas V, Koursis A, Chapple CE, Albarca Aguilera M, Meyer R et al. VarSome: the human genomic variant search engine. Bioinformatics 2019; 35: 1978–1980. [PubMed: 30376034]
13. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Leukemia. Author manuscript; available in PMC 2021 April 13.
Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17: 405–424. [PubMed: 25741868]

14. Mozzi A, Forni D, Cagliani R, Pozzoli U, Clerici M, Sironi M. Distinct selective forces and Neanderthal introgression shaped genetic diversity at genes involved in neurodevelopmental disorders. Sci Rep 2017; 7: 6116. [PubMed: 28733602]

15. Panea RI, Love CL, Shingleton JR, Reddy A, Bailey JA, Moormann AM et al. The whole genome landscape of Burkitt lymphoma subtypes. Blood 2019. doi:10.1182/blood.201901880.

16. Hill AV, Elvin J, Willis AC, Aidoo M, Allsopp CE, Gotch FM et al. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. Nature 1992; 360: 434–439. [PubMed: 1280333]

17. Liao H-M, Liu H, Lei H, Li B, Chin P-J, Tsai S et al. Frequency of EBV LMP-1 Promoter and Coding Variations in Burkitt Lymphoma Samples in Africa and South America and Peripheral Blood in Uganda. Cancers 2018; 10. doi:10.3390/cancers10060177.
Figure 1. Origin, ancestry, and relatedness of two second-degree relatives with endemic Burkitt lymphoma (eBL) in Northern Uganda.
A) Map of Uganda showing the capital city, Kampala (star with a circle), the boundary of the study area, and the location of the hospitals (red crosses) where the eBL cases were enrolled (see legend). Blue lines in the map mark all-season rivers in Uganda as a surrogate of near-homogenous high precipitation that favors high malaria transmission throughout the country. The zoom-in shows seven sub-regions (#1–7) in the study area and heterogeneity in prevalence of Plasmodium falciparum malaria measured in children aged 0–15 years during the study period (details in Supplementary data [Ref 47]). Pie charts show genomic ancestry and identity by descent (IBD) estimates. The ancestry analysis used the same methodology and African reference populations (details in Gouveia et al. 2019 [Ref 2]). B) Pathology and immunohistochemistry (IHC) of eBL tumors. Staining patterns of tumors for Case 1 and Case 2 were similar. The top two rows show: Hematoxylin and Eosin (H&E) atypical lymphocytes, CD20 positive (Dako, Carpinteria, CA, USA), CD10 positive (Novocastra, Bannockburn, IL, USA), Ki67 positive (Dako, Carpinteria, CA, USA), EBER positive (in situ hybridization [ISH] EBV (Ventana, Tuscon, AZ, USA); The bottom two rows show:
CD3 negative in atypical cells (Dako, Carpinteria, CA, USA), BCL-6 positive (Dako, Carpinteria, CA, USA), BCL-2 negative (Dako, Carpinteria, CA, USA), MYC protein positive (Epitomics, Burlingame, CA, USA); MYC translocation positive (only Case 1), fluorescent in situ hybridization probe (FISH) (Vysis LSI MYC Dual Color Break Apart Rearrangement Probe, Abbott Molecular Inc., Des Plaines, IL, USA). Images are 20X magnification. Scale bar 25um. C) Genome-wide detection of chromosomal imbalances in FFPE tumor tissue by OncoScan Array analysis. For each tumor the imbalance and B-allele frequency plots are shown. The upper panel depicts case 1 (male) in which a loss at the IGK-locus in 2p11 (left arrow), a copy-number neutral loss of 17p13.3p13.1 (including the gene TP53, middle arrow) and a heterozygous loss in 18q21.32 (including the gene CCBE1, right arrow) were called. The lower panel shows the results of case 2 with loss at the IGH-locus in 14q32 (arrow) and a probable (subclonal) gain in 1q which was just below the diagnostic threshold for calling by the evaluation pipeline.
Table 1.
Demographic, clinical, and laboratory characteristics of related Burkitt lymphoma children.

| Characteristic                     | Case 1                      | Case 2                      |
|-----------------------------------|-----------------------------|-----------------------------|
| **Demographics**                  |                             |                             |
| Season of enrollment              | Dry                         | Wet                         |
| **Clinical characteristics**      |                             |                             |
| Malaria microscopy/RDT            | Negative                    | Negative                    |
| Inpatient malaria treatment       | >=12 months ago             | Never                       |
| Outpatient malaria treatment      | >=12 months ago             | Never                       |
| Fever at enrollment              | Yes                         | Yes                         |
| Malaria fever, past 6 months     | No                          | No                          |
| Non-malaria fever, past 6 months | Yes                         | No                          |
| Number of fevers, past 12 months | Yes (>6 months ago)         | Yes (6 months ago)          |
| Number of admissions             | Once                        | None                        |
| **Laboratory results (range)**    |                             |                             |
| Hemoglobin, g/dl                  | 10-15                       | 10-15                       |
| Total white cell count, cells/10⁶ | 2.0-2.5                     | 4.0-4.5                     |
| Platelet count, cells/10⁹         | 100-150                     | 300-350                     |
| EBV Lei pattern                   | A                           | A                           |
| **Tumor characteristics**         |                             |                             |
| Anatomic sites                    | Abdomen                     | Abdomen, jaw                |
| Duration of symptoms              | 5 months                    | 3 months                    |
| Treatment                         | COM/IT-HR                   | COM/IT-HR                   |
| Number of cycles                  | 6                           | 6                           |
| Outcome                           | Cured                       | Cured                       |