White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/117358/

Version: Published Version

Article:
Johnston, W Thomas, Mutalima, Nora, Sun, David et al. (2014) Relationship between Plasmodium falciparum malaria prevalence, genetic diversity and endemic Burkitt lymphoma in Malawi. Scientific Reports. 3741. ISSN: 2045-2322

https://doi.org/10.1038/srep03741

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, as long as you credit the authors and license your new creations under the identical terms. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
B-CELL LYMPHOMA
PARASITE GENETICS

Received
7 November 2013

Accepted
20 December 2013

Published
17 January 2014

Correspondence and requests for materials should be addressed to S.M.M. (mbulaits@mail.nih.gov)

Relationship between *Plasmodium falciparum* malaria prevalence, genetic diversity and endemic Burkitt lymphoma in Malawi

W. Thomas Johnston[1], Nora Mutalima[1], David Sun[2], Benjamin Emmanuel[3], Kishor Bhatia[3], Peter Aka[3], Xiaolin Wu[2], E. Borgstein[4], G. N. Liomba[4], Steve Kamiza[4], Nyengo Mkandawire[4], Mkume Batumba[4], Lucy M. Carpenter[5], Harold Jaffe[5], Elizabeth M. Molyneux[4], James J. Goedert[3], Daniel Soppet[2], Robert Newton[1] & Sam M. Mbulaiteye[3]

[1]University of York, York, United Kingdom, [2]Frederick National Laboratory for Cancer Research, Frederick, Maryland, [3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, [4]Queen Elizabeth Hospital, Blantyre, Malawi, [5]University of Oxford, Oxford, UK.

Endemic Burkitt lymphoma (eBL) has been linked to *Plasmodium falciparum* (*Pf*) malaria infection, but the contribution of infection with multiple *Pf* genotypes is uncertain. We studied 303 eBL (cases) and 274 non eBL-related cancers (controls) in Malawi using a sensitive and specific molecular-barcode array of 24 independently segregating *Pf* single nucleotide polymorphisms. Cases had a higher *Pf* malaria prevalence than controls (64.7% versus 45.3%; odds ratio [OR] 2.1, 95% confidence interval (CI): 1.5 to 3.1). Cases and controls were similar in terms of *Pf* density (4.9 versus 4.5 log copies, p = 0.28) and having ≥3 non-clonal calls (OR 2.7, 95% CI: 0.7-9.9, P = 0.14). However, cases were more likely to have a higher *Pf* genetic diversity score (153.9 versus 133.1, p = 0.036), which measures a combination of clonal and non-clonal calls, than controls. Further work is needed to evaluate the possible role of *Pf* genetic diversity in the pathogenesis of endemic BL.

Endemic Burkitt lymphoma (eBL) is a monoclonal B-cell non-Hodgkin lymphoma that is common in equatorial Africa and Papua New Guinea, which has been linked to childhood infection with *Plasmodium falciparum* (*Pf*)[1–4], a Class 2A carcinogen for eBL[5]. Evidence for associations between eBL and *Pf* is unclear with, for example, the risk of eBL being increased in children with antibody markers of recent *Pf* infection while decreased in those with antibody markers of long-term exposure to *Pf* infection[6,7]. An alternative approach is to assess *Pf* prevalence, density, or genetic diversity as risk factors for eBL. Early studies of the association between eBL and *Pf* prevalence yielded null[8,9] or inverse associations[10], but they were limited by small sample sizes and reliance on microscopy that has variable sensitivity to detect *Pf* infection and that cannot distinguish infection with multiple *Pf* genotypes.

A recent ecological study using published data from Ghana, Uganda, and Tanzania[11], countries where *Pf* transmission intensity is moderate to high (mesoendemic to holoendemic)[12–14], showed that the age-specific risk of eBL and the average number of distinct malaria genotypes per positive blood sample both peaked between ages 5–9 years. The peaks for age-specific asymptomatic parasitaemia and parasite density, in contrast to those of eBL, both peaked at age about 2 years[12–14]. Infection with multiple *Pf* genotypes is relatively common in children in areas with holoendemic malaria[15], but its association with eBL has yet to be fully studied.

Here, we report our investigation to test the hypothesis that *Pf* prevalence, parasite density in peripheral blood, and genetic diversity are associated with eBL among 303 children with eBL (cases) compared to 274 children with non eBL-related cancers or non-malignant conditions (controls) in Malawi. *Pf* genetic diversity was measured using a sensitive and specific *Pf* molecular-barcode array[16] of 24 independently segregating *Pf* single nucleotide polymorphisms (SNPs) representative of the 3D7 *Pf* genome.

Results

Pf malaria prevalence potentially associated with eBL. Cases were similar to the controls with respect to gender, but they were slightly older than the controls (7.7 [SD 0.2] years versus 6.5 [0.3] years) (Table 1). The distribution

Table 1 | Children included in this study by cancer type showing age and sex distributions and prevalence of *P. falciparum* DNA and Epstein-Barr Virus antibody in serum

			Age						Epstein-Barr Virus antibody	
					Classes (%)					
Diagnosis	N	% Male	Mean (se)		0–5	6–10	11–15	*P. falciparum* prevalence (%)	No. Tested	Prevalence (%)
Cases	303	59.4	7.7	(0.2)	24.8	57.1	18.2	64.7	127	59.8
Controls	274	55.8	6.5	(0.3)	48.2	28.5	23.4	45.3	112	26.8
Diagnosis of controls										
Leukemias	9	22.2	8.8	(1.6)	33.3	0.0	66.7	77.8	5	20.0
Lymphomas	51	62.7	9.3	(0.5)	13.7	43.1	43.1	47.1	16	50.0
Cranial tumours	3	66.7	9.0	(1.5)	0.0	66.7	33.3	0.0	1	0.0
Neuroblastoma	15	73.3	5.5	(1.0)	60.0	33.3	6.7	53.3	5	60.0
Retinoblastoma	14	35.7	3.6	(0.6)	92.9	0.0	7.1	64.3	10	10.0
Renal Tumours	77	57.1	4.0	(0.3)	72.7	24.7	2.6	37.7	30	26.7
Hepatic Tumours	14	71.4	10.3	(0.9)	14.3	28.6	57.1	42.9	5	20.0
Bone Tumours	3	66.7	10.3	(2.0)	0.0	66.7	33.3	66.7	1	0.0
Soft Tissue Sarcomas	44	61.4	6.4	(0.6)	52.3	25.0	22.7	38.6	19	36.8
Germ Cell Tumours	16	37.5	5.9	(1.1)	50.0	31.3	18.8	50.0	8	0.0
Epithelial Tumours	6	33.3	9.0	(1.9)	16.7	16.7	66.7	33.3	1	0.0
Other Tumours	4	25.0	4.0	(1.5)	75.0	25.0	0.0	75.0	3	33.3
Non-Malignancies	18	50.0	7.0	(1.1)	38.9	33.3	27.8	50.0	8	0.0
Not well specified	31	51.6	8.1	(0.6)	19.4	61.3	19.4	35.5		

Table abbreviation: % per cent; se standard error of the mean; 8 children with Kaposi sarcoma were included in the soft tissue sarcomas; Not well specified - children didn't have enough information to classify them as a specific cancer. None of the 11 Not well specified diagnosis children who are Pf positive had enough DNA to make it the restricted dataset (see table 2).

of cases and controls across reported home districts was similar (not shown). The *Pf* prevalence as assessed by PCR analysis was 64.7% among the cases compared to 45.3% among the controls (OR = 2.1, 95% CI: 1.5–3.1). Similar results (66.9% versus 46.4%) were obtained when the analysis was restricted to a subset of 239 children that were previously tested for EBV[2] (OR = 2.9, 95%, CI: 1.6–5.4). Associations between eBL and *Pf* prevalence were evident after stratifying by EBV with eBL associated with *Pf* among 106 children with a high EBV antibody reactivity (OR = 2.3, 95% CI: 0.8–6.2) and among 133 children with negative, indeterminate or low EBV reactivity (OR = 2.5, 95% CI: 1.1–5.6).

Pf genetic diversity potentially associated with eBL. To determine whether *Pf* density was associated with eBL, we compared *Pf* log copy number per 10^5 peripheral blood mononuclear cells in cases and controls. Log copy number of *Pf* parasites was similar in eBL cases and controls (4.9 versus 4.5 log copies, p = 0.28) and it decreased, albeit, non-significantly with age in both groups. *Pf* density in younger cases and controls (0–5 years) was 4.9 log copies and in older children it was 4.7 log copies (p = 0.54).

We evaluated the association between eBL and *Pf* genetic diversity in a subset of 129 children that had samples with at least 2 *Pf* DNA copies detected and at least 20 of 24 unambiguous SNP calls, which we considered the threshold for valid results (Figure 1a). Although there was no effect of *Pf* density on the number successful SNP calls, there was a weak relationship between density and the proportion of called SNPs that were non-clonal (Figure 1b) hence *Pf* density was included as an adjustment in subsequent regression analyses. Genetic diversity at one or more of the 24 SNP locations was observed in 127 (98.5%) of the 129 children (mean number of non-clonal calls per child 11.4, standard error [se] = 0.6, Table 2). The prevalence of non-clonal calls among cases was slightly, but not significantly, increased compared to that among controls (RR = 1.3, 95% CI: 0.97–1.70, P = 0.08). The prevalence of cases with at least 3 non-clonal calls was 2.7 times (95% CI: 0.7–9.9, p = 0.14) that among controls. Mixed calls (non-concordant) may indicate presence of *Pf* variant strains at levels close to the limit of assay detection. Mixed calls were less frequent than non-clonal calls; having at least one mixed call was observed in 80 (62%) of 129 children (Table 2). Having a mixed call among cases was 3.2 times more likely in cases than in controls, but the result was not statistically significant (p = 0.18). The results were similar in an expanded subset of 160 children with at least 1 copy of *Pf* DNA (Table 2).

Graphical and spline analyses of the barcode arrays of the children ordered by either diversity score (Figure 2a, b) or by the proportion of non-clonal SNP calls (not shown) revealed a greater preponderance of cases at the more-diverse end of the *Pf* diversity scale. Corroborating this finding, cases were also found to have a higher average diversity score, which up-weighted non-clonal (scored as 10) or mixed calls (scored as 5), than controls (mean score: 153.9 [se = 5.8] versus 133.1 [se = 7.7], t-test p = 0.036) (Figure 2c).

Discussion

We report results from a case-control study of children in Malawi evaluating whether prevalence, density or genetic diversity of *Plasmodium falciparum* (*Pf*) might be the triggering malaria exposure for children with endemic Burkitt lymphoma. Associations found between *Pf* prevalence and genetic diversity with eBL agree with the well-established epidemiology of eBL, i.e., that it occurs in rural areas where *Pf* transmission is high[2,4,11,17,18]. The association between eBL and *Pf* prevalence and genetic diversity score, although modest, was robust in analyses adjusting for anti-EBV and anti-malaria antibodies, *Pf* density and in sensitivity analyses. The significant difference in *Pf* and genetic diversity score in cases and controls supports the hypothesis that genetic diversity of *Pf* may play a role in triggering the pathogenesis of eBL, which was based on a previous study showing a correlation between age-specific peaks for number of malaria genotypes and age-specific eBL peaks[11]. These results are consistent with observations that eBL is characterized by a very short doubling time (1–2 days)[18], and that the interval from initiating or promoting events to diagnosis may be comparatively short (3–8 months)[19].

Biologically, *Pf* parasites and EBV are recognized as co-factors in the genesis of eBL, but the detailed mechanisms of interaction between *Pf* parasites, the B cell compartment and EBV remain obscure. Cases in our study did not have higher average *Pf* density than controls. This contrasts with children suffering from severe acute malaria (e.g., cerebral malaria) where *Pf* density is high, but genetic diversity is low[20]. Our results perhaps suggest a different conceptual

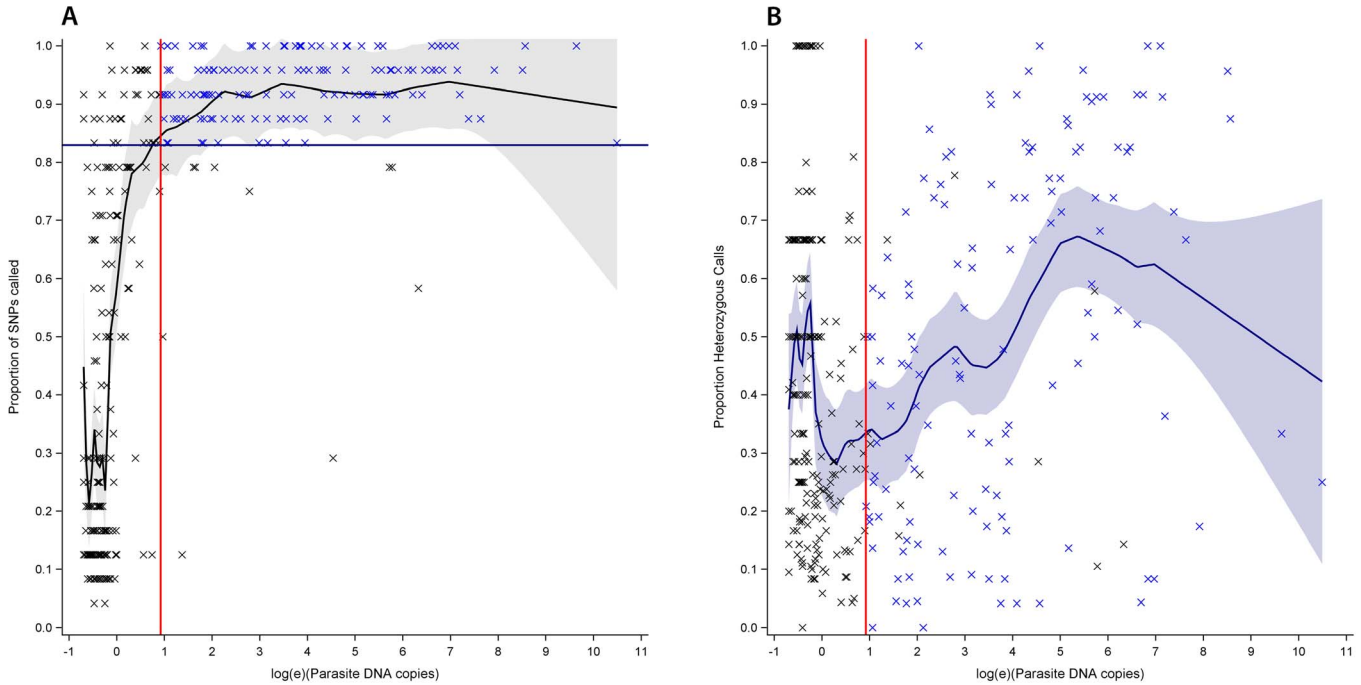

Figure 1 | The relationship between the amount of *P. falciparum* DNA isolated in cases and controls and (A) the proportion of SNP genotypes (out of 24 SNPs) determined and (B) the proportion of determined genotypes found to be non-clonal. The relationships are illustrated by loess fitted curves. Amount of DNA has been $\log_e$-transformed. Red reference line indicates 2 copies of parasite DNA present and blue reference line indicates 20 of 24 SNPs determined. Blue markers indicate those patients included in the 'high quality' subset.

model to explore with the underlying molecular mechanisms linking *Pf* density, genetic diversity, and host proteins in eBL pathogenesis. Clinical data support the notion of differences in the immunopathology of *Pf* in eBL compared to severe malaria. First, eBL is rare in children aged 0–2 years, despite being the age group when children are most vulnerable to high-density *Pf* parasitemia and severe malaria. Second, multi-clonal *Pf* infections are frequently associated with mild malaria among people with established disease immunity[21], a property that is closer to the risk profile for eBL than severe malaria.

Pf parasites modulate host defenses promoting both an immune-suppression, hyper-activation (immunosubversion)[22,23] and expansion of atypical memory B cells[24]. Possibly, subversion of immunity might be enhanced by parasite genetic diversity. If so, parasite genetic diversity could increase susceptibility of children to EBV infection or trigger reactivation of EBV among children with latent infection. Additionally, polymorphic *Pf*-encoded ligands, such as PfEMP1[25], have been shown to induce polyclonal B cell activation[26], preferentially of memory B cells (in which EBV persists and eBL develops) and to rescue tonsillar B cells from apoptosis, and to reactivate latent EBV infection[27,28]. If parasite genetic diversity enhances immune-subversion during infection, potential consequences could be: the impairment of EBV-specific T-cell response, hyper-activation of germinal centers where c-myc/Ig chromosomal rearrangements often occur, and increased survival of translocation-positive B cells. Further studies investigating the role of parasite genetic diversity based on the above points are thus needed.

Our study has some limitations. By its design, it precludes us from distinguishing whether the association precedes or follows the development of eBL[29]. The use of hospital cases and controls is a limitation because, although a similar distribution of home districts was observed for cases and controls it is not known how well this captures the actual malaria exposure related to geography. The controls were slightly younger than the cases, but this difference would bias the study towards the null, suggesting that our results may be conservative. Other limitations include a relatively small sample size, particularly in subgroup analyses. A particular strength of molecular bar code array was its design, based on 24 independently segregating SNPs scattered across the *Pf* genome, which made a direct measurement and quantification of *Pf* genetic diversity possible with a high degree of sensitivity and specificity. Despite this strength, the molecular barcode array may not be uniformly sensitive or specific for malaria clones at low quantities. Our results motivate

Table 2 | Number (%) of children with non-clonal and mixed calls in two nested subsets of the data showing the mean (se) numbers of both per child

			Non-clonal Calls						Mixed Calls			
Subset	Group	N	at least 1	(%)	at least 3	(%)	mean	(se)	at least 1	(%)	mean	(se)
At least 2 Pf DNA copies and at least 20 SNPs called	Cases	87	86	(99)	78	(90)	12.2	(0.8)	55	(63)	1.24	(0.13)
	Controls	42	41	(98)	34	(81)	9.6	(1.0)	25	(60)	1.19	(0.18)
	Total	129	127	(98)	112	(87)	11.4	(0.6)	80	(62)	1.22	(0.11)
At least 1 Pf DNA copy	Cases	105	104	(99)	91	(87)	11	(0.7)	67	(64)	1.29	(0.12)
	Controls	55	54	(98)	43	(78)	8.9	(0.9)	34	(62)	1.13	(0.15)
	Total	160	158	(99)	134	(84)	10.3	(0.5)	101	(63)	1.23	(0.10)

Table abbreviation: % per cent; se standard error of the mean; SNP single nucleotide polymorphism

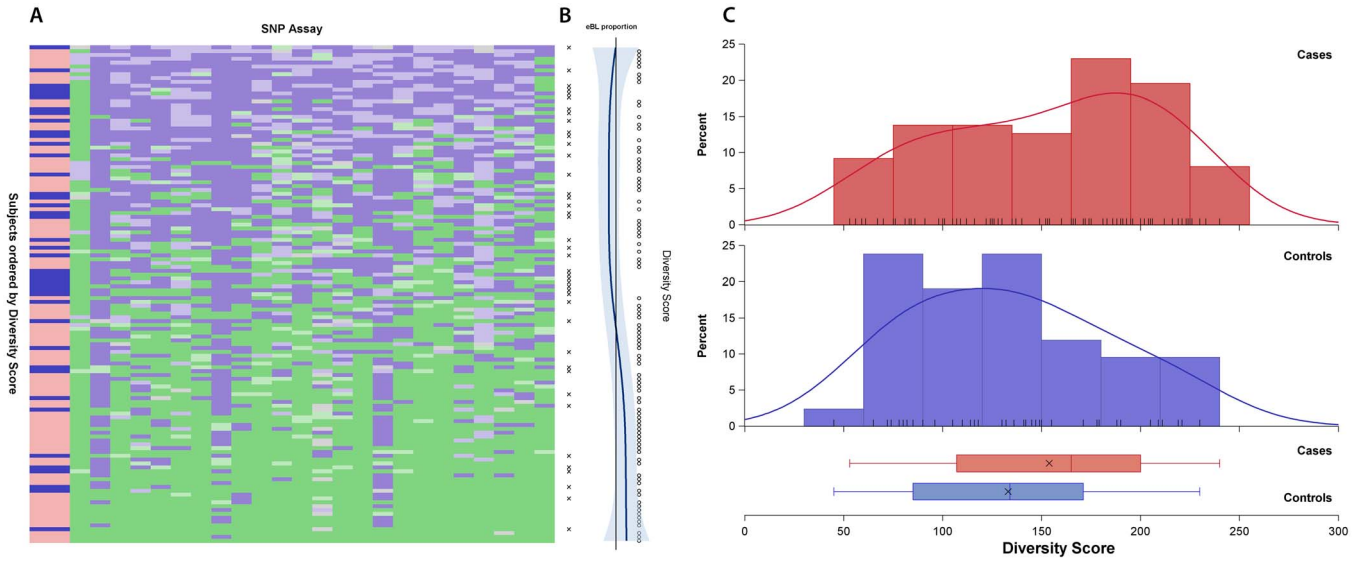

Figure 2 | The genetic diversity of *P. falciparum* isolates from 87 cases and 42 controls. A) The barcode array: The barcode for a single patient is represented in a single row whilst each column summarizes the diversity at each SNP location. Cases and controls are ordered by the diversity score (most diverse are at the bottom of the plot) and the SNPs are arranged by location in the *P. falciparum* genome – the first column indicating cancer diagnosis (Cases in red and controls in blue). SNP results are coded as follows: minor allele as lighter blue, major allele as darker blue, potentially mixed call as lighter green, non-clonal call as darker green and a failed call as light gray. B) A loess spline curve relating diversity score to the probability of being a case. The X symbols mark the rows for controls and small circles mark the rows for cases. C) A comparison of the distributions of the diversity score among cases and controls.

the adaptation of recently published malaria genome and the maturing bioinformatics computational methods to integrate genomics and proteomics[30] to investigate the role of malaria genetic diversity in carcinogenesis of eBL.

To conclude, the results of this case control study of children with endemic Burkitt lymphoma in Malawi support the hypothesis that infection with genetically diverse *Pf* parasites may be associated with eBL. It also supports the rationale of incorporating molecular methods in the study of the pathobiology of malaria in eBL. Further work is needed to evaluate the possible role, and the underlying molecular mechanisms, of *Pf* genetic diversity in the pathogenesis of eBL.

Methods

Patients and cancer diagnoses. Participants were from a case-control study of cancers in children aged 0 to 15 years conducted at the Queen Elizabeth Hospital in Blantyre, Malawi, between July 2005 and August 2010 as described elsewhere[2,3]. Cancers included Burkitt lymphoma, other haematological malignancies (leukaemia, Hodgkin lymphoma), neuroblastoma, rhabdomyosarcoma, Ewing's sarcoma, primitive neuroectodermal tumour, and Wilms' tumour. Eight children with Kaposi sarcoma were included in the current study. All cancer cases were reviewed clinically by one investigator (EM) and were confirmed by histology, cytology or other laboratory investigations when possible. Trained nurses obtained consent and administered a standardized questionnaire to the children or their parents or guardians. All children were routinely tested for HIV infection. HIV positive children were excluded from the current study. For analytic purposes, children with Burkitt lymphoma were coded as cases and children with another diagnosis as controls. The controls comprised children admitted to the same hospital with a wide range of both malignant and non-malignant conditions (Table 1).

Ethics review. The study obtained ethical approval from the Oxford Tropical Research Ethics Committee and the Malawian College of Medicine Research and Ethics Committee and exemption from ethics review by the Office of Human Subjects Research at the National Institutes of Health. All subjects gave written informed consent to participate.

DNA extraction, *P. falciparum* barcode genotyping. DNA was extracted from whole blood samples using QIAamp Blood DNA Kit (Qiagen, Inc., Valencia, CA) according to well-established protocols. Genomic DNA samples (20 ng) from blood were evaluated for Human DNA content RNAse P(ABI TaqMan 4316844, VIC) and for *Plasmodium falciparum* (*Pf*) copy number to a 519 bp segment of PF07-0076 using semi quantitative 5′nucleotidase (TaqMan)[16]. For each sample the PF assay consisted of 20 ng genomic DNA, 900 nMolar forward primer (CGACCCTGA TGTTGTTGTTGGA), 900 nMolar reverse Primer (GGCTTTTTTCCATTTCTGTA GTTAAGATTCA), 200 nMolar FAM labeled probe (CAACAGCTCCAAAATAT), 2.5 ul 2× universal master mix (Applied Biosystems), in a final volume of 5 ul. Samples were denatured at 95 degrees for 10 minutes followed by 40 cycles of amplification (95 degrees 15 sec, 60 degree 60 sec) on ABI 7900HT. Human DNA content was assessed in parallel aliquots using identical conditions and substituting primers for the Human RNAse P gene (TaqMan 4316844, VIC, product 87 bp,). The average cycle threshold of triplicate measurements for samples were compared to standards of known copy number and those samples with *P. falciparum* DNA at greater than 0.5 copies per sample were included in subsequent genotyping analysis.

Genotyping assays were performed in 96.96 dynamic arrays for SNP genotyping (SNP arrays) using the BioMark platform (Fluidigm). Each sample was assayed in quadruplicate for 24 nucleotide polymorphisms (Daniels et al, manufactured by Fluidigm). Samples comprising 20 ng genomic DNA, 50 nMolar STA primer mixture, 50 nMolar LSP Primer mixture 2.5 ul 2× universal master mix (Applied Biosystems), in a final volume of 5 ul. Samples were denatured at 95 degrees for 10 minutes followed by 15 cycles of amplification (95 degrees 15 sec, 60 degrees 120 sec) on ABI 9700. The Fluidigm SNP Array microfluidic chips were loaded with 5 ul assay comprising 2.5 μL assay loading reagent (2×) (Fluidigm 85000736), 1.0 μL 50× SNP genotyping assay mix, 7.5 uM each allelic specific primer and 20 uM locus specific primer) and 1.5 μL RNAse/DNAse free water. Samples were loaded in (5 μL) comprising 2.5 μL Biotium 2× Fast Probe master mix (Biotium 31005) 0.25 μL SNPtype sample loading reagent (20×) (Fluidigm 100-3425), 0.08 ul SNPtype Reagent (Fluidigm 100-3402), 0.03 ul ROX (Invitrogen 12223-012), 0.06 μL RNAse/DNAse free water, and 2.08 ul 5-fold dilution of the pre-amplification mixture. Control samples included a negative control (2.08 ul of water instead of genomic DNA) and positive controls of malaria genomic DNA samples (MRA-102G, MRA-150G, MRA-205G, MRA-330G) from BEI resources (Manassas, Virginia). Individual assays (5 μL) and samples (5 μL) was pipetted into separate inlets on the frame of the SNP arrays per the manufacturer instructions. Microfluidic chip loading and mixing of samples and assay mixtures in the 9216 reaction chambers of the dynamic array was carried out on the IFC Controller HX. PCR and image processing was carried out on the BioMark system (Fluidigm). Laboratory staff were masked to the case or control status of the samples.

Two issues surrounding the analyses of the SNP results were identified. Firstly, there were a number of SNP calls that were discordant when the assay was repeated: one time clonal and another non-clonal. This ambiguity might be taken as an indicator of a mixed infection where one clone is at the limit of detectability. Secondly, successful allelic *Pf* typing was defined as classification of at least 20 of the 24 SNPs on the malaria barcode at a threshold at 2 copies of *Pf* DNA in the sample. The proportion of samples with determined genotypes declined rapidly below this threshold (Figure 1a).

Statistical analyses. Initial descriptive analyses of the complete sample of 577 patients were carried out and associations between having Burkitt lymphoma or any other cancer and the presence of *P. falciparum* and/or EBV were assessed using logistic regression adjusted for age, sex and month and year of enrolment.

The main analyses of clonality were restricted to samples with at least 2 copies of *Pf* DNA and calls for at least 20 of the 24 barcode loci. Firstly, the number of non-clonal calls among cases and controls was compared using a negative binomial regression model using the number of SNP calls made as an offset (natural logarithm ($\log_e$)-transformed)) and adjusted for age, sex, year and month of enrolment and amount of *Pf* DNA present. A similar approach was used to assess the number of potentially mixed calls that occurred. Logistic regressions of the presence of at least one potentially mixed call or at least 3 non-clonal calls were also carried out. These analyses were repeated in sensitivity analyses including 160 participants with at least 1 copy of *Pf* DNA.

Finally, the unique barcode array for cases and controls was assembled and potential clustering of barcodes assessed by relating the array to cancer type status using a spline fit (with 95% confidence limits). A binary cancer type code was considered the outcome while the array was considered the predictor. The barcodes were ordered in two ways. First, a 'genetic diversity' score was assigned to each possible SNP call at each of the 24 locations: 0 for a failed call, 1 for a minor allele (<35% prevalence, as previously defined by Daniels et al[16,31]), 3 for a major allele, 5 for a potentially mixed call and 10 for a non-clonal call. The score for the entire array was obtained by summing these scores. Second, the proportion of the array with non-clonal calls was determined. For this, potentially mixed calls counted as both a clonal call and non-clonal call. Departures of the spline fit away from the proportion of patients in the sample with Burkitt lymphoma may be indicative of clusters of barcodes predominantly found in one group of patients. Two-sided p-values <0.05 were considered statistically significant; p-values between 0.1 and 0.05 were suggestive of a trend. Because this study was exploratory in nature for hypothesis generation, no adjustment was done for multiple comparisons. All analyses were undertaken with the SAS System (SAS/STAT version 12.3, SAS Institute, Cary, NC, USA).

1. Rochford, R., Cannon, M. J. & Moormann, A. M. Endemic Burkitt's lymphoma: A polymicrobial disease? *Nat Rev Microbiol* **3**, 182–187 (2005).
2. Mutalima, N. *et al.* Associations between Burkitt lymphoma among children in Malawi and infection with HIV, EBV and malaria: Results from a case-control study. *PloS One* **3**, e2505 (2008).
3. Mutalima, N. *et al.* Impact of infection with human immunodeficiency virus-1 (HIV) on the risk of cancer among children in Malawi - preliminary findings. *Infect Agent and Cancer* **5**, 5 (2010).
4. Carpenter, L. M. *et al.* Antibodies against malaria and Epstein-Barr virus in childhood Burkitt lymphoma: A case-control study in uganda. *Int J Cancer* **122**, 1319–1323 (2008).
5. Bouvard, V. *et al.* Carcinogenicity of malaria and of some polyomaviruses. *Lancet Oncol* **13**, 339–340 (2012).
6. Aka, P. *et al.* Endemic Burkitt lymphoma is associated with strength and diversity of *Plasmodium falciparum* malaria stage-specific antigen antibody response. *Blood* **122**, 629–635 (2013).
7. Guech-Ongey, M. *et al.* Antibodies reactive to *Plasmodium falciparum* serine repeat antigen in children with Burkitt lymphoma from Ghana. *Int J Cancer* **130**, 1908–1914 (2012).
8. Ziegler, J. L. *et al.* Burkitt's lymphoma and malaria. *Trans R Soc Trop Med Hyg* **66**, 285–291 (1972).
9. Feorino, P. M. & Mathews, H. M. Malaria antibody levels in patients with Burkitt's lymphoma. *Am J Trop Med Hyg* **23**, 574–576 (1974).
10. Asito, A. S. *et al.* Elevated anti-zta igg levels and EBV viral load are associated with site of tumor presentation in endemic Burkitt's lymphoma patients: A case control study. *Infect Agent and Cancer* **5**, 13 (2010).
11. Emmanuel, B. *et al.* African Burkitt lymphoma: Age-specific risk and correlations with malaria biomarkers. *Am J Trop Med Hyg* **84**, 397–401 (2011).
12. Gardiner, C., Biggar, R. J., Collins, W. E. & Nkrumah, F. K. Malaria in urban and rural areas of southern Ghana: A survey of parasitaemia, antibodies, and antimalarial practices. *Bull World Health Organ* **62**, 607–613 (1984).
13. Okello, P. E. *et al.* Variation in malaria transmission intensity in seven sites throughout Uganda. *Am J Trop Med Hyg* **75**, 219–225 (2006).
14. O'Meara, W. P., Mangeni, J. N., Steketee, R. & Greenwood, B. Changes in the burden of malaria in sub-Saharan Africa. *Lancet Infect Dis* **10**, 545–555 (2010).
15. Branch, O. H. *et al. Plasmodium falciparum* genotypes, low complexity of infection, and resistance to subsequent malaria in participants in the Asembo Bay cohort project. *Infect Immun* **69**, 7783–7792 (2001).
16. Daniels, R. *et al.* A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar J* **7**, 223 (2008).
17. Rainey, J. J. *et al.* Spatial distribution of Burkitt's lymphoma in Kenya and association with malaria risk. *Trop Med Intern Health* **12**, 936–943 (2007).
18. Iversen, U., Iversen, O. H., Bluming, A. Z., Ziegler, J. L. & Kyalwasi, S. Cell kinetics of african cases of Burkitt lymphoma. A preliminary report. *Eur J Cancer* **8**, 305–308 (1972).
19. Williams, E. H., Day, N. E. & Geser, A. G. Seasonal variation in onset of Burkitt's lymphoma in the west nile district of Uganda. *Lancet* **2**, 19–22 (1974).
20. Ofosu-Okyere, A. *et al.* Novel *Plasmodium falciparum* clones and rising clone multiplicities are associated with the increase in malaria morbidity in Ghanaian children during the transition into the high transmission season. *Parasitology* **123**, 113–123 (2001).
21. Farnert, A., Rooth, I., Svensson, Snounou, G. & Bjorkman, A. Complexity of *Plasmodium falciparum* infections is consistent over time and protects against clinical disease in tanzanian children. *J Infect Dis* **179**, 989–995 (1999).
22. Donati, D. *et al.* Increased B cell survival and preferential activation of the memory compartment by a malaria polyclonal B cell activator. *J Immunol* **177**, 3035–3044 (2006).
23. Simone, O. *et al.* TLRs innate immunereceptors and *Plasmodium falciparum* erythrocyte membrane protein 1 (pfemp1) CIDR1alpha-driven human polyclonal B-cell activation. *Acta Trop* **119**, 144–150 (2011).
24. Weiss, G. E. *et al.* Atypical memory B cells are greatly expanded in individuals living in a malaria-endemic area. *J Immunol* **183**, 2176–2182 (2009).
25. Smith, J. D. *et al.* Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
26. Donati, D. *et al.* Identification of a polyclonal B-cell activator in *Plasmodium falciparum*. *Infect Immun* **72**, 5412–5418 (2004).
27. Lam, K. M., Syed, N., Whittle, H. & Crawford, D. H. Circulating Epstein-Barr virus-carrying B cells in acute malaria. *Lancet* **337**, 876–878 (1991).
28. Chene, A. *et al.* Endemic Burkitt's lymphoma as a polymicrobial disease: New insights on the interaction between *Plasmodium falciparum* and Epstein-Barr virus. *Sem Cancer Biol* **19**, 411–420 (2009).
29. Grimes, D. A. & Schulz, K. F. Descriptive studies: What they can and cannot do. *Lancet* **359**, 145–149 (2002).
30. Crompton, P. D. *et al.* A prospective analysis of the ab response to *Plasmodium falciparum* before and after a malaria season by protein microarray. *Proc Natl Acad Sci USA* **107**, 6958–6963 (2010).
31. Milner, D. A. Jr. *et al.* Human cerebral malaria and *Plasmodium falciparum* genotypes in Malawi. *Malaria J* **11**, 35 (2012).

Acknowledgments

We are grateful to Dr. Robert J Biggar of Queensland University, Brisbane, for his comments on the manuscript. We are grateful to the participants in the study in Malawi and to the work of all the local doctors in Malawi who contributed to this work.
The work was supported by the National Cancer Institute's Director's Innovation Award to Dr. S.M. Mbulaiteye in the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute at the National Institutes of Health, Department of Health and Human Services (Support Services Contract NO2-CP-31003).

Author contributions

N.M., G.N.L., S.K., N.M., M.B. and E.M.M. conducted field work; D.S., X.W. and D.S. conducted laboratory studies; W.T.J. performed statistical analyses and prepared figures; B.N., P.A., B.K., L.C.M., H.J. and G.J.J. interpreted data; R.N. and S.M.M. conceived the idea, designed the study and supervised the work. All authors had access to data, commented on and contributed to the final draft of the manuscript. W.T.J. and N.M. Authors contributed equally to the work. R.N. is the Principal investigator of the original case-control study in Malawi.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Johnston, W.T. *et al.* Relationship between *Plasmodium falciparum* malaria prevalence, genetic diversity and endemic Burkitt lymphoma in Malawi. *Sci. Rep.* **4**, 3741; DOI:10.1038/srep03741 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0