Introduction

Multiple sclerosis (MS) is a disorder of the central nervous system that is characterized by chronic inflammation, demyelination, axonal loss and neurodegeneration [1]. Genome-wide association (GWAS) screens and meta-analyses have enabled identification of about 50 non-HLA MS risk genes [2–16]. Apart from the HLA region, the implicated genes exert modest effects at the population level with odds ratios (OR) ranging from 1.1–1.3 [17]. In a recent study, we validated the association of four risk single nucleotide polymorphisms (SNPs) with MS susceptibility [5], in a northern Spanish-Basque population, from which the CD6 SNP rs17824933 emerged with a stronger risk (OR = 1.34) [18].

CD6 is a member of the group B scavenger receptor cysteine-rich super family (SRCR-SF) [19] found on thymocytes, mature T-cells, some B-cell and natural killer (NK) subsets and is also expressed in certain parts of the brain like the cerebellum, basal ganglia, thalamus, corpus amygdaloideum, and cerebral lobi [20–23].

At the transcriptional level, in addition to the full-length form, a total of six isoforms have been reported that diversify the cytokine domains [24,25]. The ligand for CD6 is the Activated Leukocyte Cell Adhesion Molecule (ALCAM), found in the thymin epithelium and in the epithelial layer of the blood-brain barrier. ALCAM binds to the third SRCR domain of CD6 and this
interaction was recently shown to enable transmigration of CD4+ T lymphocytes across the blood-brain barrier [26].

Functional studies using mAbs to CD6 have demonstrated its role in T cell activation, proliferation [27–30] and in regulating the expression of intracellular phosphoproteins and production of proinflammatory cytokines [31]. Furthermore, CD6 is involved in the maturation of the immunological synapse (IS) by associating at the central supramolecular activation cluster (cSMAC) region [25]. However, an isoform lacking the ALCAM binding domain, CD6ΔD3 which was upregulated upon T cell activation, was not localized at the IS [25,32]. Studies using thymocytes show an increased expression of CD6 on thymocytes in the mature (single-positive) compared to the immature (double-negative, double-positive) stages and showed the negative influence of CD6 on the rate of apoptosis implicating its role in thymocyte selection [20]. The CD6ΔD3 expression was also observed to be higher on mature compared to immature thymocytes [25].

A recently published [33] correlation analysis of rs17824933 genotypes with expression of two CD6 extracellular domain isoforms (full-length and CD6ΔD3) on CD4+ and CD8+ T cells from healthy donors revealed no significant differences. However, comparison of the relative expression of the two isoforms showed the risk allele (rs17824933G) to be associated with a higher expression of the isoform lacking the ligand binding domain (CD6ΔD3).

In the present study, we performed a two-stage CD6 SNP screen to identify the putative causative variant(s) or haplotypes that contribute to increased MS susceptibility. We analyzed the effect of associated haplotypes on the cell surface expression of CD6 in T and NK cells by flow cytometry and assessed differences in proliferation and cytokine production (IFN-γ and IL-17) according to haplotype.

Materials and Methods

Sample Collections

The details of the sample collections used for the genetic study are listed in Table 1. All affected individuals meet established diagnostic criteria [34,35], and the blood samples were obtained after written informed consent from all donors, and with the approval of the institutional ethics research committees from Bilbao (Comité Ético de Investigación Clínica (CEIC) del Hospital de Basurto), San Sebastián (CEIC del Hospital de Donostia), Madrid (CEIC del Hospital Clínico San Carlos-IdISSC), Málaga (CEIC del Hospital Regional Universitario Carlos Haya), Sevilla (CEIC del Hospital Universitario Virgen Macarena), Granada (CEIC del Hospital Universitario Virgen de las Nieves), and UCSF, USA (UCSF Institutional Review Board). Initial screening was performed using the northern Spanish-Basque dataset, which includes samples collected from two centers - the Hospital de Basurto (Bilbao) and the Donostia Hospital (San Sebastián). The replication sample set included 2265 MS patients and 2600 healthy controls across the blood-brain barrier [26].

CD6 and Multiple Sclerosis Risk

The list of monoclonal antibodies (mAb) and their suppliers are provided in the Table S2. Table S3 gives the details of the combination of markers used to identify the different cell subsets. CD3 and CD56 were used to identify T (CD3+ CD56+), NK (CD3− CD56+), and NKT (CD3+ CD56+) cell subsets. Within T cells, CD4 and CD8 mark T-helper and T-cytotoxic subsets. Subsequently, CD45RA, CD27 and CD28 were used to

Statistical Analysis of Genetic Data

The data obtained was analyzed using PLINK software (version 1.07) [http://pngu.mgh.harvard.edu/purcell/plink/][37]. Hardy-Weinberg equilibrium (HWE) test was performed to check for deviations in the control population. The strength of association was assessed via the odds ratio (OR) values in the individual datasets. The Cochran-Mantel-Haenszel (CMH) test was performed on the replication and combined datasets to test for association after stratification. A test of heterogeneity (Breslow-Daytest) was performed on the replication and combined datasets to identify any heterogeneity. Haplotype analysis for the original sample set was performed using the Haplovieview software (version 4.2) [38], and the sliding window analysis and Max-T permutation P values were calculated using PLINK.

Samples for Fluorescence-activated Cell Sorting (FACS) and ELISA

10 ml fresh lithium heparinised blood was obtained by venipuncture from twenty-seven MS patients (Department of Neurology, Institute of Clinical Neurosciences, Hospital Regional Universitario Carlos Haya, Málaga, Spain) and from twelve healthy donors (University Klinikum Eppendorf, Hamburg, Germany). The clinical characteristics of these twenty-seven MS patients are listed in Table S1. Peripheral blood mononuclear cells (PBMC) were purified using a Ficoll-Hypaque gradient, as described in the supplier’s protocol (ICN Biomedicals Inc., OH, USA), counted and immediately used or cryopreserved in the presence of dimethyl sulphoxide 10% (v/v) (DMSO), and 20% fetal bovine serum and stored at −196°C until further use. When using frozen samples, PBMC were thawed immediately before use.

Flow Cytometry

The list of monoclonal antibodies (mAb) and their suppliers are provided in the Table S2. Table S3 gives the details of the combination of markers used to identify the different cell subsets. CD3 and CD56 were used to identify T (CD3+ CD56+), NK (CD3− CD56+), and NKT (CD3+ CD56+) cell subsets. Within T cells, CD4 and CD8 mark T-helper and T-cytotoxic subsets. Subsequently, CD45RA, CD27 and CD28 were used to
unstimulated, stimulated with anti-CD3 (OKT3, 100 ng/ml) with Bio-One Ltd., UK) and cultured under three different conditions: RPMI (GIBCO, Life technologies, USA) containing 5% FBS. Human NK cells can be further divided into the classical cytotoxic cells (CD45RA+CD27−CD69−CD16−), and the more immune-regulatory CD56bright (CD56+CD16+). For staining, PBMCs were thawed, washed twice, and incubated 45 min with the antibody cocktails listed in Table S3. Cells were subsequently washed, re-suspended in 300 μl FACS buffer and analyzed on a FACS Canto flow cytometer (BD Bioscience, San Diego, CA). FACS Diva software version 6.1.3 (BD Bioscience, San Diego, CA) was used to analyze the raw data.

**Cell Culture, Proliferation and ELISA Assays for Cytokine Measurement**

PBMCs (10⁶ cells) from each donor were suspended in 500 μl of RPMI (GIBCO, Life technologies, USA) containing 5% FBS (Biochrom AG, Germany), aliquoted into a 48-well plate (Greiner Bio-One Ltd., UK) and cultured under three different conditions: unstimulated, stimulated with anti-CD3 (OKT3, 100 ng/ml) with or without anti-CD6 mAb (161.8, concentration 10 μg/ml) (kindly provided by Prof. Francisco Lozano, IDIBAPS, Facultat de Medicina, Universitat de Barcelona, Barcelona). To determine cell proliferation, PBMC from healthy donors were labeled with mEFluor670 (eBioscience) according to the manufacturer’s protocol. In brief, eFluor stock reconstitution and dilutions were done in PBS at 4°C and incubation with the dye was performed at room temperature (RT) for ten minutes. To stop labeling, standard medium was added and incubated on ice for five minutes. Labeled cells were then transferred into 48-well plates with the respective stimuli for further cell culture. Cells were counterstained with specific lineage markers to assess proliferation at day 0 and 3 by assessing the percentage of cells that had undergone division determined by flow cytometry. The supernatants from the culture were collected on the third day and quantified for the cytokines IL-17 and IFN-γ using Ready-Set-Go ELISA kits (eBiosciences, San Diego, CA, USA). The cytokine concentration (pg/ml) was calculated using the standard curves generated using the respective standards.

**Statistical Analysis of FACS and ELISA Data**

ANOVA and two-tailed Student’s t test were used to calculate the differences in surface expression, while differences in proliferation and cytokine production between the different haplotypes were assessed by the non-parametric Mann-Whitney U test using Graphpad Prism software version 5 (GraphPad Software, La Jolla, CA).

**Results**

A CD6 Haplotype Containing Two Non-synonymous SNPs in the 2nd SRCR Domain is Associated with MS

The first stage of the study was performed on a dataset from northern Spain (Basque Country) constituting a total of 825 MS patients and 814 healthy controls. Nine haplotype tagging (haptag) SNPs (Hapmap release 27) were selected to cover CD6 gene variability. Three further non-synonymous SNPs were included – rs11230563 (R225W) and rs2074225 (A257V) in the second SRCR domain (Exon 4) and rs12360861 (A271T) in the third SRCR domain (Exon 5). In addition, rs17824933 [5,18] and rs11230563 (R225W) was also included as it demonstrated a trend towards association (P = 0.08, OR = 1.13) in the first-stage screen. Three SNPs emerged with nominal significance from the

| Table 1. Clinical characteristics of the MS patients included in the genetic study. |
|---------------------------------------------|
| **NORTHERN SPANISH-BASQUE** | **REPLICATION DATASET** |
| BILBAO | SAN SEBASTIÁN | UCSF WHITES | ANDALUCÍA | MADRID |
| Controls | Cases | Controls | Cases | Controls | Cases | Controls | Cases | Controls | Cases |
| Participants | | | | | | | | |
| Total number | 565 | 573 | 249 | 250 | 450 | 507 | 1340 | 1119 | 810 | 639 |
| % Male/Females | 31.2/68.8 | 26.5/73.5 | 34/66 | 38/62 | 32.9/67.1 | 32/68 | 27.3/62.7 | 26.8/62.5 | 45.4/53.2 | 34.7/62.8 |
| Disease course (%) | | | | | | | | |
| RR & SP/PP/PR – 87.6/10/0 – 88/2/0/ – 79.7/3.1/0.4/0.5/15.9 | 83.9/7/0.1 | | | | | | | |
| /other/no details /1.6/0.8 | 0.8/9.2 | | | | | | | |
| Age at Onset | | | | | | | | |
| Mean ± S.D. – | 30.5±10.10 | – | 31.6±10.3 – | 33.3±9.3 – | 29.4±9.9 – | 28.9±8.9 |
| EDSS | | | | | | | | |
| Mean ± S.D. – | 3.08±2.1 | – | 3.91±2.6 – | 2.04±1.96 | – | 3.2±1.8 – | 2.9±2.3 |

**Abbreviations:** RR = Relapsing-remitting, SP = Secondary progressive, PP = Primary progressive, PR = Primary relapsing, S.D. = Standard deviation.

doi:10.1371/journal.pone.0062376.t001
replication set; rs17824933, rs11230559 and rs2074225 (Table 3). Combined analysis of all datasets showed association of rs11230559, rs2074225, rs17824933 and rs650258.

Analysis of LD patterns revealed that rs11230559 is in strong LD with rs17824933 SNP in the northern Spanish dataset (D' = 0.99, r² = 0.93; Figure 2), and also in the combined dataset (not shown), in agreement with data from the 1000 Genomes Project (D' = 1, r² = 0.85; Figure 2). Both these SNPs are in strong LD with two non-synonymous SNPs – rs11230562 (T217M) in the SRCR domain 2 (Exon 4) (D' = 1, r² = 0.94 with rs17824933 and D' = 1, r² = 0.8 with rs11230559) and rs2074233 (G606S) in the cytoplasmic domain (Exon 11) (D' = 1, r² = 0.84 with rs17824933 and D' = 1, r² = 0.898 with rs11230559). However, rs650258 showed limited or no LD with any of the three non-synonymous SNPs rs11230562 (D' = 0.224, r² = 0.038), rs2074225

Figure 1. Locations of the 14 Stage 1 SNPs and their linkage disequilibrium (LD) patterns at the CD6 locus. (A) LD block images with the approximate locations of the 14 SNP markers chosen for this study, recombination spots and LD patterns within the CEU LD plot (Hapmap, version 27) of the CD6 gene, and (B) LD patterns in the northern Spanish-Basque population using the confidence interval method (Gabriel et al. [54]; generated using the Haplovew software).

doi:10.1371/journal.pone.0062376.g001

Table 2. Association analysis of 14 CD6 SNPs in the northern Spanish-Basque dataset.

| SNP ID | POWER(1) (%) | BP | LOCATION | AA CHANGE | MINOR/MAJOR ALLELE (2) | HWE P-VALUE | FREQ. (3) (4) | P-VALUE OR (95% CI) (3) |
|--------|--------------|----|----------|-----------|------------------------|-------------|---------------|-----------------------|
| rs3019561 | 96          | 60500429 | Intron 1 | T/G       |            | 0.93/0.92 | 0.25/0.24 | 0.44 | 1.07 (0.91 – 1.25) |
| rs3019562 | 99          | 60501644 | Intron 1 | C/G       |            | 0.67/0.32 | 0.51/0.47 | 0.04 | 1.16 (1.01 – 1.33) |
| rs3019548 | 97          | 60505515 | Intron 1 | G/A       |            | 0.83/0.17 | 0.40/0.38 | 0.16 | 1.11 (0.96 – 1.28) |
| rs2905506 | 95          | 60506624 | Intron 1 | T/C       |            | 0.72/1 | 0.26/0.25 | 0.36 | 1.08 (0.92 – 1.26) |
| rs11230548 | 95        | 60508145 | Intron 1 | A/C       |            | 0.9/0.55 | 0.86/0.82 | 0.008 | 1.28 (1.06 – 1.56) |
| rs17824933 | 94        | 60517188 | Intron 1 | G/C       |            | 0.73/0.27 | 0.29/0.25 | 0.007 | 1.24 (1.06 – 1.45) |
| rs11230555 | 92        | 60519710 | Intron 1 | A/C       |            | 0.84/0.16 | 0.78/0.77 | 0.57 | 1.05 (0.89 – 1.25) |
| rs916811 | 97          | 60520408 | Intron 1 | A/G       |            | 0.46/0.54 | 0.79/0.72 | 3.3 × 10⁻⁵ | 1.40 (1.21 – 1.63) |
| rs11230559 | 95          | 60526110 | Intron 1 | C/T       |            | 0.87/0.13 | 0.30/0.25 | 0.004 | 1.25 (1.07 – 1.46) |
| rs2237997 | 98          | 60528666 | Intron 1 | C/T       |            | 0.30/0.70 | 0.38/0.35 | 0.07 | 1.14 (0.99 – 1.32) |
| rs11230563 | 98          | 60532785 | Exon 4 | R225W      | C/T       | 0.12/0.82 | 0.40/0.36 | 0.08 | 1.13 (0.98 – 1.31) |
| rs2074225 | 99          | 60532882 | Exon 4 | A257V      | C/T       | 0.55/0.45 | 0.71/0.63 | 3.1 × 10⁻⁵ | 1.40 (1.21 – 1.63) |
| rs12360861 | 75          | 60533649 | Exon 5 | A271T      | A/G       | 0.36/0.64 | 0.81/0.81 | 0.79 | 1.03 (0.85 – 1.21) |
| rs650258(5) | –            | 60588858 | 3‘ Intergenic | T/C       |            | 0.2/0.06 | 0.56/0.51 | 0.0052 | 1.23 (1.06 – 1.41) |

(1) Power was calculated with an OR = 1.34 based on the original OR for rs17824933 in the Spanish-Basque dataset [18] using the CEU frequencies (NCBI) for each SNP.
(2) The risk alleles are underlined.
(3) Values represented in terms of the risk allele.
(4) Genotyping success rates were above 95% for all the SNPs typed.
(5) rs650258 was included based on the association observed in a recent genome-wide screen in MS [17].

doi:10.1371/journal.pone.0062376.t002
Haplotype analysis was performed on all datasets. A sliding window analysis using two, three and four markers was done followed by 10K max(T) permutation analysis. Furthermore, two- and three-marker analysis using at least one of the two non-synonymous SNPs was also performed (Table S4). The sliding window analysis showed stronger association (OMNI-BUS) of combinations including both the exonic non-synonymous SNPs rs11230563 and rs2074225 (P_{max(T)} permutation = 1.610^{-4}), and of rs2074225 and rs650258 (P_{max(T)} permutation = 1.610^{-4}) (Figure 3, Table S4). Addition of rs17824933 or rs11230559 to the haplotype markers rs11230563-rs2074225 did not significantly alter the haplotype frequencies (Table S5) or increase the strength of association (data not shown), indicating that the association was in essence explained by the haplotype rs11230563-rs2074225 (P_{max(T)} permutation = 1.610^{-4}) (Figure 3, Table S4). The use of aggressive tagging on two-marker and three-marker haplotypes showed that rs11230563 could be tagged by markers rs916811 and/or rs11230559 along with rs2074225 (r^2 = 0.946) and also by rs916811 and/or rs17824933 with rs2074225 (r^2 = 0.892). Similarly, the SNP rs2074225 was

### Table 3. Replication and combined analysis of the most significant SNPs.

| SNP | Replication | Combined |
|-----|-------------|----------|
| rs11230548 (A/C) | 0.64 | 0.25 |
| rs17824933 (G/C) | 0.02 | 0.44 |
| rs916811 (A/G) | 1.00 | 0.49 |
| rs11230559 (C/T) | 0.02 | 0.20 |
| rs11230563 (T/C) | 0.77 | 0.69 |
| rs2074225 (C/T) | 0.03 | 0.53 |
| rs650258 (T/C) | 0.08 | 0.46 |

Abbreviations: CMH = Cochran-Mantel-Haenszel test, BD = Breslow-Day test, OR = Odds ratio, CI = Confidence interval.

(1) The risk alleles (with respect to the Basque dataset) are underlined.

(2) The ORs are represented for the risk alleles found in the first-stage screen.

(3) rs11230563 was included because it is a non-synonymous SNP that substitutes R225W, and showed a trend towards association in the first stage screen.

(4) rs650258 was included based on the association observed in the Basque dataset.

doi:10.1371/journal.pone.0062376.t003

Figure 2. LD values (D'/r^2 values) between each of the 14 SNP markers genotyped in the Spanish-Basque population. The figure represents a comparison of the LD values between the 14 SNPs genotyped in the northern Spanish-Basque population (D'/r^2 values) using the values from the CEU 1000 genomes project and from the data generated from the Spanish-Basque population (823 cases/814 controls). The values in white are the D'/r^2 values from the CEU 1000 Genomes Project, while the blue-shaded ones are generated for the northern Spanish-Basque dataset using Haplovie software (version 4.2).

doi:10.1371/journal.pone.0062376.g002

\[ D' = 0.095, r^2 = 0.005 \] and \( r' = 0.149 \) (http://www.broadinstitute.org/mpg/snap/ldsearch.php) [42].

\[ (D' = 0.095, r^2 = 0.005) \] and \( (D' = 1, r^2 = 0.149) \) (Figure 2) showed a trend towards association in the first stage screen.

\( rs11230563 \) was included because it is a non-synonymous SNP that substitutes R225W, and showed a trend towards association in the first stage screen.

\( rs650258 \) was included based on the association observed in the Basque dataset.

\( D' = 0.095, r^2 = 0.005 \) and \( r' = 0.149 \) (http://www.broadinstitute.org/mpg/snap/ldsearch.php) [42].

\( (D' = 1, r^2 = 0.149) \) (Figure 2) showed a trend towards association in the first stage screen.

\( rs11230563 \) was included because it is a non-synonymous SNP that substitutes R225W, and showed a trend towards association in the first stage screen.
subsets according to their naïve, memory or effector phenotypes, CD4 SNPs, which was confirmed by logistic regression analysis. A strong association of the two domain-2 (Exon-4) non-synonymous susceptibility to MS. Sliding window haplotype analysis revealed independent effects exerted by the SNPs rs2074225 and rs2074225, and conditioning rs1230563 with rs2074225 showed strong additive effects with $P = 6.2 \times 10^{-7}$ (Table S6).

In summary, we identified two novel SNPs associated with MS risk, that is, rs1230559 located in the intronic region, and rs2074225, a non-synonymous SNP located in the 2nd SRCR domain, and validated the association of rs650258 [17] with susceptibility to MS. Sliding window haplotype analysis revealed strong association of the two domain-2 (Exon-4) non-synonymous SNPs, which was confirmed by logistic regression analysis.

The rs11230563-rs2074225 CD6 Haplotype Modifies CD6 Expression on CD4+ and CD8+ Naïve T Cells

FACS analysis was performed on lymphocytes of MS patients genotyped for both CD6 markers SNPs rs11230563 and rs2074225. CD6 was expressed at very high levels on all CD4+ T cell subsets, and to a lesser extent on CD8+ and NKT cells (Figure 4, Figure S2). In these three lymphocyte subsets, CD6 expression levels varied significantly, – with the protective haplotype (rs11230563-rs2074225) CC expressing higher CD6 levels than both the risk haplotype CT and the TT haplotype that provides mild risk (Figure 4). These differences were prominent among the CD4+ ($P_{ANOVA} = 0.0008$) and CD8+ ($P_{ANOVA} = 0.001$) lymphocytes, but were also observed in NKT cells ($P_{ANOVA} = 0.02$). Upon separation of the CD4+ and CD8+ cell subsets according to their naïve, memory or effector phenotypes, we found the greater differences in the naïve (CD4+ naïve cells, $P_{ANOVA} = 0.0001$; CD8+ naïve cells, $P_{ANOVA} < 0.0001$), and the CD4+ central memory subset ($P_{ANOVA} = 0.01$), while no differences were found in the effector subsets and in the terminally differentiated effector cells (CD27+ CD28+) (Figure 4). Of note, NKT cells also expressed CD6, and the haplotype-related differences were also observed in this population ($P_{ANOVA} = 0.02$). NK cells expressed much lower levels of CD6 on their surface, and no significant changes in expression were observed in relation to the CD6 haplotypes. Similar observations were made when the non-synonymous SNPs were analyzed individually (Figure S3, S4).

Haplotype-specific Differences in Proliferation and Cytokine Secretion upon T-cell Stimulation

The differential expression of CD6 from the different genotypes and haplotypes points to a possible effect on T cell function. Further experiments were aimed at identifying differences in proliferation and in cytokine production. Polyclonally stimulated T cells were assessed to determining whether co-stimulation of CD6 using specific antibodies had a haplotype-related effect on proliferation and on cytokine production. For this purpose, PBMCs were stimulated with anti-CD3 in the presence or absence of anti-CD6 161.8 mAb. Our data showed no significant differences in proliferation and in the levels of IFN-γ and IL-17A secretion between the haplotypes or different conditions (Figure S5). However, even though not significant, donors with the CC haplotype showed relatively higher proliferation and higher levels of IL-17A than those with the CT haplotype. Furthermore, co-stimulation with the anti-CD6 mAb, 161.8 showed a mild decrease in proliferation and IL-17A production (Figure S5).

Discussion

In this study, we aimed to identify the most important MS-associated CD6 SNP variants through a fine-mapping exercise. Our data showed a strong association of a non-synonymous SNP rs2074225 with susceptibility in the Spanish-Basque dataset that was replicated in the combined dataset (Table 2, 3). Furthermore, the analysis also revealed association of rs17824933 and rs11230559, both of which are in strong LD with each other ($r^2>0.8$) and with two other non-synonymous SNPs located in exon 4 (rs11230562) and exon 11 (rs2074223). The association of these SNPs was observed to be stronger in the original dataset (Spanish-Basque residents) when compared to the replication sets; and this may be related to factors such as disease heterogeneity, and the relatively high degree of genetic homogeneity in the Spanish Basque geographic sampling area [43].

Haplotype analysis using sliding window and two-marker approaches implicated a role of two non-synonymous SNPs rs11230563 (R225W) and rs2074225 (A257V). We analyzed
expression changes of CD6 on different immune cell types according to genotypes for the individual non-synonymous SNPs genotypes or their haplotypes. Our screen included, apart from the CD4+ and CD8+ T-cells studied by Kofler et al., [33], also NK cells, NK1.1 cells, and T-effector cell subsets in an attempt to identify the cell type more strongly influenced by the CD6 non-synonymous SNP genotype/haplotype. Since the mAb used in this study for FACS targets the SRCR domain 1 region, discrimination between different isoforms was not possible. Our data showed a more pronounced effect of the rs11230563-rs2074225 haplotype compared to the individual non-synonymous SNPs (Figure 4, Figure S3, S4). The expression trends showed the protective haplotype (CC) to yield higher surface expression of CD6 when compared to the risk haplotypes (CT and TT), which is in agreement with both the trends observed by Kofler et al. on the full length isoform [33] and the analysis of Heap et al., [44] who reported allele imbalances of eight SNPs in the CD6 region including rs2074225, rs11230562, and - as inferred though LD patterns, rs17824933.

We also observed that the expression of the TT (mild risk) haplotype cells was lower than that from the CT (risk) haplotype. While the CT haplotype is intermediate in terms of CD6 expression between the protective (CC) and mild risk (TT) haplotype in the CD4+ and CD8+ T cells, its expression shows a non-significant trend to be higher in CD4+ and CD8+ effector memory T cells. Thus, our data suggests that this CD6-conferring genetic risk is complex and may result in changes in both CD6 expression levels and ligand binding or signal transduction via the non-synonymous SRCR domain 2 SNPs rs11230563 (R225W) and rs2074225 (A257V) in addition to the cytoplasmic domain SNP rs2074232 (G606S).

The contribution of the isoform lacking the ligand-binding domain (CD6ΔD3), of which expression is inversely correlated with that of the full-length form [33], should be factored into the elucidation of CD6 function. Kofler et al., [33] observed no significant differences in CD6 expression of the full length form between genotypes, but reported increased relative expression of CD6ΔD3 in both CD4+ and CD8+ T-cells in individuals homozygous for the risk allele rs17824933GG. In our study, the naive CD4+ and CD8+ T-cell populations were more strongly affected by the CD6 haplotype than effector and central memory cells. This is important as CD6 is highly expressed in thymic cells and is thought to aid positive selection and provide resistance to apoptosis [20]. Furthermore, Singer et al. [20] and Castro et al. [25] showed the two isoforms to be more highly expressed on mature cells. In the thymus, CD6ΔD3 is under-expressed in the double-positive cells, while the expression of the full-length form is favored which contributes to their survival.

In addition, being a member of the scavenger family, both the isoforms of CD6 appear to bind to bacteria, and may act therefore in an early infectious phase putatively associated with onset of MS, even if the precise mechanism is as yet unknown. Binding studies with LPS and anti-CD6 mAbs showed phosphorylation of ERK1/2 and p38 MAPK indicating activation of the MAPK pathway [45,46]. Observation of proliferation and cytokine secretion patterns showed a trend towards reduced proliferation by the risk haplotype when compared to the protective haplotype. These findings are in league with the findings of Kofler et al. [33], who observed reduced cell proliferation among the donors with the risk genotype rs17824933GG. Similar trends to that of proliferation were observed with IL-17, where higher levels of this cytokine were observed with the protective haplotype (CC) than with the risk haplotype (CT) (not-significant trend; Figure S5). Since addition of anti-CD6 did not significantly alter the proliferation rate and the production of cytokine, it could be inferred that the anti-CD6 mAb could lead to generalized blocking of proliferation and cytokine production and is not SNP-dependent. The IL-17 trends, though inconclusive, do suggest that allelic variation in CD6 may be associated with altered cytokine secretion from Th17 cells. Given the capability of CD6 to bind to bacteria/LPS [45], and the evidence for a role of IL-17A in mediating protection against various pathogens [47,48], this may be of relevance to the concept of infectious agents acting as triggers for MS. Alternatively, observation of higher IL-17A levels among those with the protective haplotype could indicate a protective/ anti-inflammatory role of IL-17A. The ability of IL-17A to protect against development of autoimmune uveitis and ulcerative colitis [49,50] has been demonstrated; however, this contrasts with the reported disease-promoting role of Th17 cells in MS [51–53], and will need further clarification.

Taken together, our results indicate that non-synonymous polymorphic variations in the 2nd SRCR domain are associated with functional changes. Screening of the different cell types showed the most significant expression differences in CD4+ and CD8+ naive cells, suggesting that phenotypic expression of CD6 variation may affect the early stages of cell-mediated immune responses.

**Supporting Information**

**Figure S1** Plots showing Sequenom-based clustering of the alleles from the dataset of Bilbao. Each of the axes represents an allele and each sample in the graph is represented as a dot. Samples homozygous for any of the alleles fall near the x or y-axis, while the heterozygotes lie in the graph area between the two axes. (TIF)

**Figure S2** FACS-gating strategy. PBMCs were stained with anti-CD3, CD56, CD16, CD4, CD8 and CD6 antibodies. After gating on lymphocytes, the T and NK cell subsets were defined as shown in the top left panel. From the NK cells, NKbright and NKdim cells were identified according to the level of CD56 and the presence of CD16. T-helper and T cytotoxic lymphocytes were identified by CD4 and CD8 staining, respectively. In these two populations, CD45RA and CD27 were used to define naive, effector and memory cell subsets, as indicated in the two top right panels. The lower panels show the corresponding histograms depicting the expression of CD6 in each of the aforementioned subsets. NK histogram, solid line indicates NKbright and dashed...
Figure S3 Comparison of CD6 expression on the different cell types with respect to rs11230563 genotypes. (TIF)

Figure S4 Comparison of CD6 expression on the different cell types with respect to rs2074225 genotypes. (TIF)

Figure S5 Comparison of proliferation and cytokine levels (pg/ml) between the three haplotypes. PBMCs representing each of the haplotypes (CC = protective, CT = high risk, TT = mild risk) from the healthy donors were cultured in a 48-well plate for three different stimulatory conditions – unstimulated, OKT3 stimulated (100 ng/ml) with/out anti-CD6 161.B (10 μg/ml). Proliferation was assessed by measuring 10^6 cells stained with eFluor670 on the pad software (version 5).

Table S1 Demographic and clinical variables of the PBMC samples collected from MS patients for the functional study. (DOC)

Table S2 Details of the monoclonal antibodies used in flow cytometry. (DOC)

Table S3 The combination of monoclonal antibodies used for cell surface staining. (DOC)

Table S4 OMNIBUS association P-values of haplotypes generated using sliding window or multiple marker analysis on the merged datasets. (DOC)

Table S5 Comparison of the haplotype frequencies in the four datasets of (A) the 2 associated NS haplotype SNP markers, or (B) three NS SNP markers. (DOC)

Table S6 Logistic regression analysis on the combined dataset showing additive effects of the SNPs when conditioned to rs11230563, rs2074225 and rs650258. (DOC)

Acknowledgments

We wish to express our gratitude to all MS patients who have contributed to this study via provision of biological samples.

Author Contributions

Conceived and designed the experiments: BS AC ET A. Antigué J. Performed the experiments: BS AC FM A. Alcina JRO MF TDO EU. Analyzed the data: BS AC IA ET KV. Contributed reagents/materials/analysis tools: MGB IA OF ML JO MJPM JRO A. Antigué RA. Wrote the paper: BS ET FM EU JRO KV.
29. Bott CM, Doshi JB, Morimoto C, Romain PL, Fox DA (1993) Activation of human T cells through CD6: functional effects of a novel anti-CD6 monoclonal antibody and definition of four epitopes of the CD6 glycoprotein. Int Immunol. 5: 783–792.

30. Oliveira MI, Gonçalves CM, Pinto M, Fabre S, Santos AM, et al. (2011) CD6 attenuates early and late signaling events, setting thresholds for T-cell activation. Eur J Immunol. 42: 195–205.

31. Hassan NJ, Barclay AN, Brown MH (2004) Optimal T cell activation requires the engagement of CD6 and CD166. Eur J Immunol. 34: 930–940.

32. Gimferrer I, Calvo M, Mittelbrunn M, Farnós M, Sarrias MR, et al. (2004) Relevance of CD6-mediated interactions in T cell activation and proliferation. J Immunol. 173: 2282–2270.

33. Kotler DM, Severson CA, Mousissian N, De Jager PL, Hafler DA (2011) The CD6 Multiple Sclerosis Susceptibility Allele Is Associated with Alterations in CD4+ T Cell Proliferation. J Immunol. 187: 3286–3291.

34. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, et al. (1983) New Diagnostic Criteria for Multiple Sclerosis: Guidelines for Research Protocols. Ann Neurol. 13: 227–231.

35. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, et al. (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 50: 121–127.

36. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 38: 209–213.

37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81: 559–575.

38. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

39. Hamann D, Baars PA, Rep MH, Hooshrav K, Kreithof-Garde SR, et al. (1997) Phenotypic and functional separation of memory and effector human CD8+ T cells. J Exp Med. 186: 1407–1418.

40. Koch S, Larbi A, Derhovanessian E, Özcelik D, Naumova E, et al. (2008) Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. Immun Ageing 5: 6.

41. Lanier LL, Le AM, Cevc G, Loken MR, Phillips JH, et al. (1986) The relationship of CD16 (Leu-11) and Leu-19 (NKH-1) antigen expression on human peripheral blood NK cells and cytotoxic T lymphocytes. J Immunol. 136: 4480–4486.

42. Johnson AD, Handsaker RE, Palti SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics (Oxford, England) 24: 2938–2939.

43. Rodríguez-Espinosa N, Alvarez-Busto J, Imaz I, Regueiro M, Azcárate MN, et al. (2010) High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. Human genetics 128: 113–117.

44. Heap GA, Yang JH, Downes K, Healy BC, Hunt KA, et al. (2010) Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome sequencing. Hum Mol Genet. 19: 122–134.

45. Sarrias MR, Farnós M, Mota R, Sánchez-Barbero F, Báezé A, et al. (2007) CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock. Proc Natl Acad Sci USA 104: 11724–11729.

46. Báezé A, Sarrias MR, Farnós M, Gimferrer I, Serra-Page C, et al. (2006) Mitogen-activated protein kinase pathway activation by the CD6 lymphocyte surface receptor. J Immunol. 177: 1132–1159.

47. Waite JC, Skokos D (2012) Th17 response and inflammatory autoimmune diseases. International journal of inflammation 2012: 10.

48. Curtis MM, Way SS (2009) Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens. Immunology 126: 177–185.

49. Ke Y, Liu K, Huang G-Q, Cui Y, Kaplan HJ, et al. (2009) Anti-inflammatory role of IL-17 in experimental autoimmune uveitis. Journal of immunology (Baltimore, Md: 1950) 182: 3183–3190.

50. O'Connor W, Kamanaka M, Booth CJ, Tovan T, Nakae S, et al. (2009) A protective function for interleukin 17A in T cell-mediated intestinal inflammation. Nature immunology 10: 603–609.

51. Brucklacher-Waldert V, Stuerner K, Kolster M, Wolthausen J, Tolosa E (2009) Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis. Brain 132: 3329–3341.

52. Durelli L, Coni I, Clerico M, Borelli D, Contessa G, et al. (2009) T-helper 17 cells expand in multiple sclerosis and are inhibited by interferon-beta. Ann Neurol. 65: 499–509.

53. Kebir H, Kreyenberg K, Ergan I, Dodelet-Devillers A, Cayrol R, et al. (2007) Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. Nat Med. 13: 1173–1175.

54. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. Science 296: 2225–2229.