A novel class of somatic mutations in blood detected preferentially in CD8+ cells

Miko Valori a, Lilja Jansson a, Anna Kiviharju a, Pekka Ellonen b, Hanna Rajala c, Shady Adnan Awad c,d, Satu Mustjoki c,d, Pentti J. Tienari a,*

ABSTRACT

Somatic mutations have a central role in cancer but their role in other diseases such as autoimmune disorders is poorly understood. Earlier work has provided indirect evidence of rare somatic mutations in autoreactive T-lymphocytes in multiple sclerosis (MS) patients but such mutations have not been identified thus far. We analysed somatic mutations in blood in 16 patients with relapsing MS and 4 with other neurological autoimmune disease. To facilitate the detection of somatic mutations CD4+, CD8+, CD19+ and CD4−/CD8− T-lymphocytes were separated. We performed next-generation DNA sequencing targeting 986 immune-related genes. Somatic mutations were called by comparing the sequence data of each cell subpopulation to other subpopulations of the same patient and validated by amplicon sequencing. We found non-synonymous somatic mutations in 12 (60%) patients (10 MS, 1 myasthenia gravis, 1 narcolepsy). There were 27 mutations, all different and mostly novel (67%). They were discovered at subpopulation-wise allelic fractions of 0.2%–6.6% (median 0.95%). Multiple mutations were found in 8 patients. The mutations were enriched in CD8+ cells (85% of mutations). In follow-up after a median time of 2.3 years, 96% of the mutations were still detectable. These results unravel a novel class of persistent somatic mutations, many of which were in genes that may play a role in autoimmunity (ATM, BTK, CD46, CD180, CLIP2, HMMR, IKZF3, ITGB3, KIR3DL2, MAPK10, CD56/NCAM1, RBM6, RORA, RPA1 and STAT3). Whether some of this class of mutations plays a role in disease is currently unclear, but these results define an interesting hitherto unknown research target for future studies.

1. Introduction

Somatic mutations are non-inherited alterations in DNA that occur in a progenitor cell and become transmitted to its descendant population (clone). They arise at a low frequency by chance, typically during DNA replication, and can be caused by environmental factors such as chemicals or ionizing radiation. Virtually all cancer cells have such acquired mutations and several types of somatic mutations in blood leukocytes have been characterized in e.g., leukemias [1].

Besides patients with hematological cancers, relatively little is known on the frequency and nature of somatic mutations in blood leukocytes. Age-related somatic mutations that confer growth advantage have been recently detected in blood in association with clonal hematopoiesis [2–5]. Two large-scale studies have shed light to the landscape of blood cell somatic mutations occurring in population cohorts not selected for cancer or hematologic phenotypes [3,4]. Both studies tested whole blood DNA and focused on mutations occurring with a large allelic fraction of ca. 10–20% in whole exome sequence data [3] or in a set of 160 genes implicated in hematologic cancers [4]. Somatic mutations were found in approximately 1% of subjects under age 50, and in 10% of subjects older than 70 years. Both studies identified mutational hotspots, DNMT3A, TET2, and ASXL mutations constituted >2/3 of the mutations (2, 3).

There are only few studies on leukocyte somatic mutations in autoimmune disease (other than somatic hypermutation of immunoglobulin genes). Holzelova et al. [6] reported somatic mutations in FAS in a fraction of CD4+ and CD8+ T-lymphocytes in children with an autoimmune lymphoproliferative syndrome. Similar rare autoimmune diseases have been described in conjunction with somatic mutations in the NRAS and KRAS genes [7]. Whether somatic mutations have a role in more common autoimmune disorders has not been established. A recent study by us identified somatic mutations in the STAT3 gene in CD8+ T-cells in 40% of patients with large granular lymphocytic...
leukemia [8]. Interestingly, STAT3 mutation positive patients presented with rheumatoid arthritis significantly more often than mutation negative patients, suggesting a possible role of STAT3 mutations in these autoimmune symptoms.

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system and among the most common causes of neurological disability in young adults. The cause of MS is not known, but it is assumed to be an autoimmune disorder. Multiple lines of evidence suggest that at least relapsing forms of MS would be driven by blood leukocyte dysfunction. >100 genetic loci are known to be linked with MS predisposition, the overwhelming majority of which are in genes or regions active in leukocytes [9,10]. The most potent drugs for relapsing MS (natalizumab, alemtuzumab, fingolimod, daclizumab) are mainly targeted against lymphocytes circulating in the blood [11]. Moreover, Epstein-Barr virus (EBV) is a well-documented environmental risk factor of both MS as well as lymphoid and epithelial malignancies, the risk of MS being especially high in association with EBV infectious mononucleosis [12].

Against this background, it is of interest to analyze blood leukocyte somatic mutations in MS. Indirect evidence of somatic mutations has been previously reported in MS by using the hypoxanthine guanine phosphoribosyltransferase (HPRT) assay [13–15]. This assay measures 6-thioguanine resistance of cultured cells, caused by inactivating somatic mutations of the HPRT gene or other mechanisms. The estimated mutation frequency in T-lymphocytes is approximately 5 × 10⁻⁶ in adults [14]. Allegretta et al. reported on the presence of HPRT-deficient T-lymphocyte clones in MS and, interestingly, some of these mutant clones, but not wild-type clones, were potentially autoreactive, i.e. proliferated in response to a myelin autoantigen [13]. Another study reported higher HPRT mutant frequency in MS patients’ T lymphocytes as compared to controls [15]. A pilot study on direct detection of somatic mutations in MS has been reported [16]. The study was focused on single CD4 + T-lymphocytes, derived from two MS patients’ cerebrospinal fluid. In an exome sequencing analysis of 21 individual whole genome amplified cells thousands of variants were reported, but the authors concluded that they were unable to confidently identify true somatic mutations, as these could not be discerned from whole genome amplification induced mutational noise. The study points to technical obstacles when mutations in small cell populations are screened.

In the present study we analysed whether nonsynonymous somatic mutations can be reliably found in the blood. We enriched peripheral blood mononuclear cells (PBMCs) into subpopulations and utilized next-generation sequencing of 986 immune-related genes (with additional validation sequencing), allowing the detection of mutations with low allelic fractions (down to 0.1–0.2%). We addressed the following questions. Are somatic mutations detectable with this approach? What is their frequency among the patients and allelic fraction within cell populations? Do the mutations preferentially occur in a particular cell population? Are there mutational hotspots? This information would provide important insights into the design of studies in larger materials and reveal details about the general landscape of blood somatic mutations in non-cancerous settings.

2. Results

2.1. Screening for somatic mutations in peripheral blood mononuclear cell populations

We obtained blood samples from 20 patients. 16 of the patients had relapsing MS, 2 had narcolepsy, one had myasthenia gravis, and one had stiff person syndrome and systemic lupus erythematosus (clinical details given in Table 1). We used immunomagnetic beads to enrich PBMCs into CD4 +, CD8 + and CD19 + cell populations and a remaining CD4−CD8−CD19− triple negative cell population. These markers were chosen to capture basic subgroups of interest.

We used high-depth next generation sequencing (NGS) to screen the DNAs for somatic mutations. The sequencing covered all exons of a panel of 986 immune-related genes (Table S1). We restricted our sequencing efforts to this predefined panel (referred to as Immunopanel-986 from hereon) in order to cost-effectively approach a target sequencing depth of up to 1000×. In total we successfully sequenced 71 cell populations obtained from the PBMCs of the 20 patients (mean depths within coding regions: CD4 + 753x, CD8 + 723x, CD19 + 671x, CD4−CD8−CD19− 491x). From this Immunopanel-986 sequencing data, we performed somatic variant calling to find variants that were detectable in a particular cell population and were not present in other (control) cell populations of the same patient (more details in Materials and Methods, Screening phase data analysis). We were able to produce a list of 154 putative somatic mutations for further validation using another sequencing method (amplicon sequencing).

2.2. Amplicon sequencing confirms 27 nonsynonymous somatic mutations

Using DNA from the same cell populations in which the putative somatic variants were called, we prepared validation PCR amplicons corresponding to the 154 putative nonsynonymous somatic variant calls. We again employed NGS and sequenced the patient DNA amplicons along with amplicons prepared from control DNA to an excessive mean depth of over 100,000×. We were able validate 27 of the 154 variant calls as true somatic mutations. Both screening and validation step sequencing depth and allelic fractions for each confirmed mutation are given in Table S2, along with p-values for the somatic mutation calls.

A subset of 8 mutations were tested in two PCR replicates from whole blood DNA. All mutations were detectable in whole blood DNA in both replicates. The mean allelic fraction of these 8 mutations was 0.35% in whole blood as compared to 2.25% (6.4 fold higher) in the enriched subpopulation of cells (Supplementary Table S3).

A technical reason for the inclusion of so many false positives (n = 127; 82%) was found to be counting some sequence mismatches twice, when the two opposite direction reads of paired end sequencing overlapped on short DNA fragments. This inflated the read count for some rare observations. In a post hoc analysis, we observed that counting the mismatches present on such overlapping paired end reads only once per pair eliminated the false positives, although with a cost in sensitivity (missing 16 of the 27 true somatic variants).

2.3. Somatic mutation frequencies

Out of the 20 patients, we were able to validate somatic mutations in 12 (60%). Eight of the 12 patients had multiple mutations (2–4 per patient, Table 2). Ten of the patients with somatic mutations had MS, while the other two had myasthenia gravis and narcolepsy.

The highest allelic fraction observed for a somatic mutation was 4.6%, which would correspond to a clone size of 9.2% assuming a heterozygous zygous mutation. The lowest allelic fraction of a somatic mutation was 0.35% in whole blood as compared to 2.25% (6.4 fold higher) in the enriched subpopulation of cells (Supplementary Table S3).

A technical reason for the inclusion of so many false positives (n = 127; 82%) was found to be counting some sequence mismatches twice, when the two opposite direction reads of paired end sequencing overlapped on short DNA fragments. This inflated the read count for some rare observations. In a post hoc analysis, we observed that counting the mismatches present on such overlapping paired end reads only once per pair eliminated the false positives, although with a cost in sensitivity (missing 16 of the 27 true somatic variants).

2.4. Unique and predominantly novel mutations in each patient

Table 3 shows the 27 validated somatic mutations. In total, we observed 23 single nucleotide variants (SNV) that were predicted to result in an amino acid changes, 2 in-frame 3-bp deletions, one splice-site SNV, and one premature stop-codon producing SNV. Each patient had a unique mutation profile – there were no mutational hotspots. Two patients had mutations in the CD56/NCAM1 gene, but the mutations were at different sites.

Out of the 27 somatic mutations, 19 (67%) were novel, i.e. they were not found in COSMIC [17], dbSNP [18] or ExAC [19] databases, whereas 8 of the somatic mutations could be found in at least one of these databases. Nineteen (78%) of the mutations were predicted deleterious by the CADD [20] algorithm (Table 3).
We observed no association between patient’s age (p = 0.23, two-tailed t-test) or disease duration (p = 0.93) and mutation positivity in this dataset. Normal expression of the genes, in which mutations were detected was analysed by RNA sequencing in reference cell in this dataset. Normal expression of the genes, in which mutations were observed in the CD8+ cell population. Although we sequenced a similar amount of CD4+ and CD19+ cells (based on sequencing depths), we were not able to validate more than a single mutation (including one, which was not anymore detectable), while one remained at the same level. The increases and decreases were mostly 50% or less in magnitude (in 19 of the 27 mutations). Table S6 shows the detailed changes in somatic mutation allelic fractions over time. One mutated CD8+ clone (carrying a deletion in the IKZF3 gene) increased 7-fold in allelic fraction from 2.3% to 16.4%. This MS patient had changed treatment from baseline natalizumab, which induced lymphocytosis (blood lymphocyte count was 4.09 × 10^6/ml). This data suggest that fingolimod targets other CD8+ populations more efficiently than the mutant clone. P-values and related observation counts for persistence data are listed in Table S7.

2.5. Enrichment of somatic mutations in the CD8+ cell population

Intriguingly, the clear majority (n = 23, 85%) of the 27 somatic mutations were observed in the CD8+ cell population. Although we sequenced a similar amount of CD4+ and CD19+ cells (based on sequencing depths), we were not able to validate more than a single somatic mutation in each of these cell populations (Table 4).

For the CD8+ and CD4+ cell populations, we performed T cell receptor \( V_{\beta} \) fluorescence-activated cell sorting (FACS) analysis to characterize clone sizes in the patient samples. Ten patients (50%) had an expanded “large” CD8+ clone (see definition in Table S5 footnote), but the presence of a large clone did not significantly predict the detection of a somatic mutation (p = 0.65, Table S5). Hence, preselection of the samples by T cell receptor \( V_{\beta} \) FACS does not seem to facilitate the detection of this type of mutant clones with low allelic fractions. A representative T cell receptor \( V_{\beta} \) FACS analysis of one patient (MS-8) is given in Supplementary Fig. S1. The \( V_{\beta} \) quantifications of all 20 patients are given in Supplementary Fig. S2.

2.6. Persistence of detected somatic mutations

We collected a second blood sample from each of the patients whose initial sample contained one or more somatic mutations (median sampling interval 2.3 years). We separated the cell populations from PBMCs and again performed high depth amplicon sequencing in order to examine whether the somatic mutations persisted. All but one (26 of 27) of the somatic mutations were detected in the second sample. There was no general trend towards increase or decrease in allelic fraction: 14 mutations increased and 12 decreased in allelic fraction (including one, which was not anymore detectable), while one remained at the same level. The increases and decreases were mostly 50% or less in magnitude (in 19 of the 27 mutations). Table S6 shows the detailed changes in somatic mutation allelic fractions over time. One mutated CD8+ clone (carrying a deletion in the IKZF3 gene) increased 7-fold in allelic fraction from 2.3% to 16.4%. This MS patient had changed treatment from baseline natalizumab, which induced lymphocytosis (blood lymphocyte count was 4.09 × 10^6/ml) to fingolimod, which induced lymphopenia (blood lymphocyte count was 0.66 × 10^6/ml). This data suggest that fingolimod targets other CD8+ populations more efficiently than the mutant clone. P-values and related observation counts for persistence data are listed in Table S7.

---

**Table 1**

Baseline characteristics of the patients.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Diagnosis</th>
<th>Immunological medication</th>
<th>Age</th>
<th>Sex</th>
<th>Disease duration (years)</th>
<th>Separated cell populations a</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-1</td>
<td>Multiple sclerosis</td>
<td>Natalizumab</td>
<td>35</td>
<td>M</td>
<td>2</td>
<td>CD4+, CD8+, CD19+</td>
</tr>
<tr>
<td>MS-2</td>
<td>Multiple sclerosis</td>
<td>None b</td>
<td>55</td>
<td>F</td>
<td>9</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-3</td>
<td>Multiple sclerosis</td>
<td>Natalizumab</td>
<td>38</td>
<td>M</td>
<td>4</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-4</td>
<td>Multiple sclerosis</td>
<td>Natalizumab</td>
<td>32</td>
<td>F</td>
<td>3</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MG-5</td>
<td>Myasthenia gravis</td>
<td>Ritusimab, mycophenolate mofetil, gamaglobulin</td>
<td>67</td>
<td>F</td>
<td>0.5</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>SP-6</td>
<td>Stiff-person syndrome (GAD antibody positive), SLE</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
<td>36</td>
<td>F</td>
<td>13</td>
<td>CD4+, CD8+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-7</td>
<td>Multiple sclerosis</td>
<td>Glatiramer acetate</td>
<td>57</td>
<td>F</td>
<td>13</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-8</td>
<td>Multiple sclerosis</td>
<td>Azathioprine</td>
<td>59</td>
<td>F</td>
<td>35</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>NL-9</td>
<td>Narcolepsy (type 2 non-cataplectic)</td>
<td>None</td>
<td>24</td>
<td>M</td>
<td>3</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-10</td>
<td>Multiple sclerosis</td>
<td>Interferon beta</td>
<td>25</td>
<td>F</td>
<td>2</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-11</td>
<td>Multiple sclerosis + type 1 diabetes</td>
<td>None</td>
<td>40</td>
<td>F</td>
<td>6</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-12</td>
<td>Multiple sclerosis</td>
<td>Interferon beta</td>
<td>46</td>
<td>F</td>
<td>0.5</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-14</td>
<td>Multiple sclerosis</td>
<td>None</td>
<td>48</td>
<td>F</td>
<td>7</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>NL-16</td>
<td>Narcolepsy-cataplexy (type-2)</td>
<td>None</td>
<td>22</td>
<td>M</td>
<td>3</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-17</td>
<td>Multiple sclerosis</td>
<td>Natalizumab</td>
<td>58</td>
<td>F</td>
<td>15</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-19</td>
<td>Multiple sclerosis</td>
<td>Glatiramer acetate</td>
<td>35</td>
<td>F</td>
<td>16</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-21</td>
<td>Multiple sclerosis</td>
<td>Interferon beta</td>
<td>54</td>
<td>F</td>
<td>10</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-22</td>
<td>Multiple sclerosis</td>
<td>None</td>
<td>43</td>
<td>F</td>
<td>0.5</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-23</td>
<td>Multiple sclerosis</td>
<td>None</td>
<td>29</td>
<td>F</td>
<td>0.5</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
<tr>
<td>MS-24</td>
<td>Multiple sclerosis</td>
<td>Interferon beta</td>
<td>22</td>
<td>F</td>
<td>1.5</td>
<td>CD4+, CD8+, CD19+, CD4−CD8−CD19−</td>
</tr>
</tbody>
</table>

a Cell populations with adequate amount of DNA for sequencing are shown.
b Previous treatment with interferon-beta, glatiramer acetate, natalizumab, mitoxanthrone, and azathioprine.

---

**Table 2**

Number of somatic mutations discovered per patient.

<table>
<thead>
<tr>
<th>N mutations</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>MS-8, MS-19, MS-21</td>
</tr>
<tr>
<td>3</td>
<td>MS-2</td>
</tr>
<tr>
<td>2</td>
<td>MS-1, MS-3, NL-9, MS-12</td>
</tr>
<tr>
<td>1</td>
<td>MG-5, MS-14, MS-22, MS-23</td>
</tr>
<tr>
<td>0</td>
<td>MS-4, SP-6, MS-7, MS-10, MS-11, NL-16, MS-17, MS-24</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Distribution of the allelic fractions of validated somatic mutations.
The allelic fraction of a somatic mutation is calculated as the percentage of sequencing reads that contain the variant base. Genomic coordinates are shown in Table S2. Sample Gene AA change CADD score
4.63% 4.71% MS-12-CD8 + CD1C R89C 10.0
4.03% 2.71% MS-8-CD19 + TRAF2 409–410 LF/LF 26.8
2.88% 2.29% MS-19-C19 + CD46 A250Ts 23.4
2.81% 2.70% MS-21-CD8 + RBM6 P245s 25.3
2.71% 2.70% MS-2-CD8 + A2ML1s R1001W 32.0
2.28% 1.62% MS-1-CD19 + IKZF3 155–156 FT/FT 22.9
2.23% 2.14% MS-2-CD8 + BTK Splicingb 24.2
1.58% 1.69% MS-19-CD19 + PTPMT1 H165V 17.0
1.58% 1.29% MG-5-others CD56/NCAM1 G417Hbs 34.0
1.36% 0.70% MS-19-CD19 + ITGAL Q581K 16.4
1.22% 0.86% MS-21-CD8 + CD56/NCAM1 R158Xsb 18.2
1.21% 1.68% MS-8-CD8 + RPA1 L394P 25.5
0.98% 1.32% MS-8-CD8 + KIR3DL2 D339Y 23.4
0.95% 1.15% MS-21-CD8 + RORA R533Q 23.0
0.83% 1.23% MS-19-CD19 + P501s S237t 12.4
0.83% 1.04% MS-3-CD8 + HMMR E405Ksb 26.5
0.80% 1.31% NL-9-CD8 + C0s A298Hsb 23.1
0.68% 0.80% NL-9-CD8 + CLIP2 R932H 34.0
0.59% 1.07% MS-1-CD8 + MBL2s A175s 8.9
0.46% 0.74% MS-2-CD4 + CD180 D526Y 24.4
0.45% 0.70% MS-21-CD8 + STAT3 D661Y 34.0
0.40% 0.74% MS-22-CD8 + INSR A119Vb 31.0
0.35% 0.46% MS-14-CD8 + ITGB3 F229Lsb 29.7
0.32% 0.53% MS-8-CD8 + TLR7s N275Kb 0.000
0.23% 0.60% MS-3-CD8 + MAP3K10s S136b 22.2
0.19% 1.34% MS-12-others ATM L2519Vsb 26.3
0.19% 0.56% MS-23-CD8 + CFH V407L 0.001

The table under 0.5% for most of the mutations. We con

b Indicates a novel mutation.
c CADD scores of 20 or more can be considered predictably deleterious.

c3

c4
A notable observation was the presence of the STAT3*D661Y

2.7. Autoimmunity associated STAT3*D661Y mutation in CD8 + cells of an MS patient

A notable observation was the presence of the STAT3*D661Y

The mean sequencing depth at sites where validated somatic muta-

tions were detected in the screening phase was 1062× (median 934×)
after duplicate read removal (Table S8). Therefore the target sequencing
depth should be set to at least 1000× in subsequent studies, with added
depth beyond this expected to generate more findings. With more
target genes/regions and more sequencing depth it is likely that more
mutations would have been discovered in the present dataset. It should
be noted that because of sequencing noise the depth cannot be
increased indefinitely. The highest quality base calls that the Illumina
platform produces have a quality score estimate of around 1/10,000
errors, which was found to be empirically quite accurate in our data.
In order to find somatic mutations as rare as 1/10,000 and beyond,
specialized techniques such as tagging DNA fragments with random barcodes [24] would have to be employed.

A striking finding in our study was the clear enrichment (85%) of the
somatic mutations in the CD8 + cell population. This enrichment may stem from technical reasons (detectability) or reflect overall higher mutational load in CD8 + cells. CD8 + memory cells have been shown to produce larger clones than other populations [25] rendering muta-
tions more readily detectable. On the other hand, mutagenesis itself may be especially prominent in CD8 + cells. CD8 + cells exhibit more clonal expansions upon antigen stimulus than CD4 + cells [26] and especially large clonal expansions of CD8 cells have been demonstrated in infectious mononucleosis [27]. More mitoses generate more mutations. Whether other biological reasons are also involved in the enrichment of mutation in CD8 cells (e.g. higher sensitivity to external mutagens, lower fidelity of DNA repair, lower rate of mutant clone elimination) is currently less clear. From an autoimmunity

Table 4 Number of validated somatic mutations per cell population.

<table>
<thead>
<tr>
<th>Cell population type</th>
<th>Mutations detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 +</td>
<td>1</td>
</tr>
<tr>
<td>CD8 +</td>
<td>23</td>
</tr>
<tr>
<td>CD19 +</td>
<td>1</td>
</tr>
<tr>
<td>CD4 – CD8 – CD19 –</td>
<td>2</td>
</tr>
</tbody>
</table>
CD19+ cells. Purities were measured according to the manufacturer’s protocol. In addition to the CD4+, CD8+, and CD19+ populations, we also utilized the CD28 antibody. From the separated cell populations, DNA and RNA were extracted using the InviTrap Spin Universal RNA Mini Kit (Stratec Biomedical, Birkenfeld, Germany) according to manufacturer’s instructions. The purities of the separated CD4+, CD8+, and CD19+ cells were >90% for the CD4+ and CD8+ cells and >70% for the CD19+ cells.

4.2. Cell separation and DNA extraction

Mononuclear cells were extracted from 120 to 140 ml venous EDTA blood using Ficoll-Paque PLUS (GE Healthcare Bio-Sciences, Piscataway, NJ, USA). First, 13 ml of Ficoll-Paque was added to a centrifuge tube. Then, 9 ml of blood diluted with 28 ml of PBS was layered on top of it. The tube was centrifuged at 800 × g for 30 min after which the mononuclear cell layer was transferred to a new tube with a pipette. The cells were washed twice, using PBS and centrifugation at 500 × g for 10 min. From the mononuclear cells, nucleic acids were extracted using an OctoMACS magnetic separator (Miltenyi Biotec, Bergisch Gladbach, Germany) following the manufacturer’s protocol. In addition to the CD4+, CD8+, and CD19+ fractions, a CD4−CD8−CD19− fraction was separated by using all of the above bead types to give a negative selection. Cultivation numbers for the used beads were 130-050-301 (CD19), 130-045-201 (CD8) and 130-045-101 (CD4). One separated CD8+ fraction was cultured with a CD28 antibody. From the separated cell populations, DNA and RNA were extracted using the InviTrap Spin Universal RNA Mini Kit (Stratec Biomedical, Birkenfeld, Germany) according to manufacturer’s instructions. The purities of the separated CD4+, CD8+, and CD19+ cells were tested in a subsample by flow cytometric analysis, in which the observed purities were >90% for the CD4+ and CD8+ cells and >70% for the CD19+ cells.

4.3. Custom gene panel

A custom designed gene panel referred to as Immunopanel-986 was utilized for the mutation screening step. The panel consists of 986 genes related to immunity and cancer, the bulk of which originated from the curated InnateDB database [32] with additional targets such as genes related to the JAK-STAT pathway compiled from different publications (for full gene list, see Table S1).

4.4. Screening phase library preparation and sequencing

DNA from the separated cell populations was fragmented using a Covaris S2 instrument (Covaris, Woburn, MA, USA) and then a sequencing library was prepared according to the NEBNext DNA Sample Prep Master Mix Set 1 (New England BioLabs, Ipswich, MA, USA) manual with some minor exceptions. Target enrichment for all exons of the genes included in the Immunopanel-986 gene panel was achieved with the NimbleGen SeqCap exon capture system (Roche NimbleGen, Madison, WI, USA). Library Purification was performed using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The target enriched library was sequenced with a HiSeq 2500 instrument (Illumina, San Diego, CA, USA) using paired end 101 bp reads.

4.5. Screening phase data analysis

Sequencing reads were mapped against the GRCh37 reference genome using BWA-MEM [33] afteradapter trimming with the Trimomatic [34] software. PCR duplicates were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard). VarScan2 [35] was used for somatic mutation calling. Each sequenced cell population acted in turn as the VarScan2 “tumor” (target) sample from which somatic mutations were called, and merged data from the other cell populations originating from the same patient acted as the VarScan2 “normal” (control) sample. The variant detection strategy of VarScan2 was augmented with custom logic to enable defining a cutoff for somatic variants of high enough confidence to be included in a validation step. The custom steps were as follows: first, all mismatches against the reference genome were tested using a binomial test with error rate estimated from the data to discard random non-systematic sequencing noise, Bonferroni corrected by the number of tests (size of target area * 4, a test for each possible base). The variant sites were filtered for strand bias and against possible contamination from germline variants from other samples. Then the target (“tumor”) and control (“normal”) samples were compared using Fisher’s exact test to show somatic status for sites that passed the first filters. The putative somatic variant list from all samples was then FDR corrected using the Benjamini and Hochberg method with FDR cutoff set to 0.05 to yield a final list of variants to feed into validation. Variants were annotated using the ANNOVAR [36] package. Execution of the analysis pipeline was performed using the Snakemake [37] software.

4.6. Validation and re-analysis of mutations by amplicon sequencing

PCR amplicons for each of the putative somatic variants identified in the screening phase were prepared following the “Illumina 16S Metagenomic Sequencing Library Preparation” [38] protocol, adjusted for targets other than the bacterial 16s gene. Briefly, all designed PCR primer pairs had their forward primer prepended with the sequence TCGTGGCACGCTACATGTTAGAAAAACAG and the reverse primer prepended with the sequence GTCTCGTGGGCTCGGAGATGTGTATAAGACAG. Primer oligos were ordered from Sigma Aldrich (St. Louis, MO, USA) and PCR was carried out using the KAPA HiFi polymerase (Kapa Biosystems, Wilmington, MA, USA). A second round of PCR utilizing the overhang sequences was then performed to incorporate full Illumina sequencing adapters using primers from the Nextera XT Index Kit (Illumina, San Diego, CA, USA). The amplicons were purified after both
PCR steps using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The resulting library was sequenced using a MiSeq instrument (Illumina, San Diego, CA, USA) using 200 bp paired end reads. Amplicon sequencing was performed using DNA from the same cell populations in which a putative somatic variant was identified. In addition, amplicons of all targets were prepared using separate donor control DNA. Eight variants that were validated in the first sequencing run were also tested in whole blood DNA using the same protocol. Sequencing reads were mapped against the GRCh37 reference genome using BWA-MEM [33] after adapter trimming with the Trimmomatic [34] software. VarScan2 [35] was used in validation mode to report somatic p-values (Fisher’s exact test) at each of the genomic coordinates containing a variant to be validated. Data from patient cell populations served as the VarScan “tumor” input and data from the control DNA amplicons as the VarScan “normal” input. The data was also inspected directly from samtools [39] mpileup output. The somatic variants from Immunopanel HiSeq data were considered successfully verified if they met a p-value threshold of 0.001 in this data, Bonferroni corrected by the number of verified targets.

4.7. FACS

An analysis of T cell receptor Vβ (beta chain variable region) repertoire was performed in CD4+ and CD8+ cells. The IO Test Beta Mark TCR Vβ Repertoire Kit (Beckman Coulter, Brea, CA, USA) was used and flow cytometry was performed with a BD FACSaria II instrument (BD Biosciences, San Jose, CA, USA). A cell population was considered to contain a large clonal expansion if a Vβ type was seen in 10% or more of the cells and was present at over twice of the intensity of a population control frequency.

4.8. RNA sequencing

Gene expression in the target cell subpopulations (CD4+, CD8+, CD19+ and CD4+/CD8–/CD19−) was analysed by RNA sequencing of two MS patients: MS17 and one patient (MS-0), who was not included to the somatic mutation analysis (female, no treatment, age 42). Additionally, CD8+ target cell basal RNA expression was analysed in two patients (MS-7, MS-8) and in one patient (MS-8) after stimulation with CD3 and CD28 antibodies [40]. Sequencing libraries were prepared using a strand-specific kit (New England Biolabs, Ipswich, MA, USA) and sequenced with the MiSeq instrument (Illumina, San Diego, CA, USA) with at least 25 million 300 bp read pairs produced per library. Reads were aligned using TopHat2 [41] and expression quantified using HTSeq-Count [42] and a custom script. The expression data is given in Table S4. Genes were considered expressed, when the cut-off 1 was reached in the FPKM value.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.clim.2016.11.018.

Acknowledgements

This study has been supported by the Helsinki University Hospital (grants TYH2013407 and TYH 2014248 for PJT; TYH 2014234 for SM), Biogen Finland (grant 3756 for PJT), Multiple Sclerosis Foundation of Finland (2013 for PJT, 2016 for MV), European Research Council (project M-IMM for SM) Academy of Finland (grant 134150 for PJT, 136181 for SM), the Finnish Cancer Institute (SM), and University of Helsinki. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources and Dr. Nina Peitsaro and Noora Aarnio in the Biomedicum Helsinki Flow cytometry core facility for FACS services. We thank Mr. Harri Salo, Dr. Jarno Honkanen and Prof. Outi Vaarala for CD82 stimulation assay. We thank Prof. Kimmo Pörkka for discussions in the study design.

References


Acknowledgements

This study has been supported by the Helsinki University Hospital (grants TYH2013407 and TYH 2014248 for PJT; TYH 2014234 for SM), Biogen Finland (grant 3756 for PJT), Multiple Sclerosis Foundation of Finland (2013 for PJT, 2016 for MV), European Research Council (project M-IMM for SM) Academy of Finland (grant 134150 for PJT, 136181 for SM), the Finnish Cancer Institute (SM), and University of Helsinki. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources and Dr. Nina Peitsaro and Noora Aarnio in the Biomedicum Helsinki Flow cytometry core facility for FACS services. We thank Mr. Harri Salo, Dr. Jarno Honkanen and Prof. Outi Vaarala for CD82 stimulation assay. We thank Prof. Kimmo Pörkka for discussions in the study design.

References


