Cellular immune endophenotypes separating early and late-onset myasthenia gravis

Jakob Theorell[1,2]*, Nicolas Ruffin[3], Andrew Fower[4], Chiara Sorini[3], Philip Ambrose[4], Valentina Damato[5], Lahiru Handunnetthi[6,7], Isabel Leite[4], Sarosh R. Irani[4,8], Susanna Brauner[2,3], Adam E. Handel[4], Fredrik Piehl[2,3]

1. Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden
2. Department of Neurology, Karolinska University Hospital, Stockholm, Sweden
3. Department of Clinical Neuroscience, Karolinska Institutet, and Center for Molecular Medicine, Karolinska University Hospital Stockholm, Sweden.
4. Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
5. Department of Neuroscience and Psychology, University of Florence, Florence, Italy
6. Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
7. Department Psychiatry, University of Oxford, Oxford, United Kingdom
8. Department of Neuroscience, Mayo Clinic Florida, Jacksonville, Florida, USA

Correspondence to: Jakob.theorell@ki.se, Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden; Mailing address: CIM/ANA Futura, Alfred Nobles Allé 8B, 141 52 Huddinge, Karolinska Institutet

Conflict of interest: see specific supplementary attachment for all financial relationships for all authors. This work was funded by academic funding bodies.

Abstract

The two main subgroups of autoimmune myasthenia gravis, a neuromuscular junction disorder associated with muscle weakness, are the early and late-onset forms, defined by onset before or after 50 years of age. Both carry acetylcholine-receptor autoantibodies, but differ in sex ratios, genetics and occurrence of disease-specific thymus inflammation. By applying multimodal techniques, including deep spectral cytometric phenotyping and single cell sequencing to peripheral blood and thymic lymphocyte samples we explored the possibility to discriminate the two forms by cellular immune phenotyping. Analyzing two independent cohorts we identified distinct immunological differences driven by three main lymphocyte populations. Lower frequencies of mucosa-associated invariant T cells and naïve CD8 T cells were observed in late-onset myasthenia, suggesting enhanced immune senescence. Further, a highly differentiated, canonical natural killer cell population was reduced in early-onset myasthenia, which was negatively correlated with the degree of thymic inflammation. Using only the frequency of these three populations, correct myasthenia subgroup assignment could be predicted with an accuracy of 90%. The NK cell population negatively associated to early-onset disease had a similar association to thymic hyperlasia, whereas the two T-cell populations point to enhanced immune senescence in late-onset myasthenia gravis. These distinct immunocellular endophenotypes for early- and late onset disease suggest differences in the immunopathogenic processes. Together with demographic factors and other disease subgroup-specific features, the frequency of the identified cell subpopulations may improve clinical classification, in turn of relevance for channeling to interventions.

Introduction

Myasthenia Gravis (MG) is an autoimmune neurological disease caused by autoantibodies targeting the neuromuscular junction, resulting in skeletal muscle weakness and sometimes life-threatening respiratory crises. A minor proportion only have ocular symptoms, whereas the majority (80-85%) suffer from generalized MG involving multiple muscle groups. (1) (2) Approximately 80% of patients have autoantibodies against the acetylcholine receptor (AChR) and the remaining carry autoantibodies targeting other neuromuscular junction proteins, or are denoted as seronegative.(3)

Patients with AChR autoantibodies (AChR+ MG) can be subdivided either into Early-onset or Late-onset MG (EOMG and LOMG), based on disease onset prior to or after the age of 50 years, or the less common thymoma-associated MG with underlying paraneoplastic pathophysiology. (4) (5) (6) EOMG and LOMG display contrasting features, implying at least partially different pathogenetic mechanisms; i) >70% females in EOMG vs <40% in LOMG (4); ii) carriers of HLA-B*08:01 are at risk of developing EOMG, whereas no HLA gene confers a marked risk specifically for the development LOMG in the latest genome-wide association study meta-analysis,(7); iii) multiple autoantibody specificities may be observed in LOMG (e.g., anti-striated muscle antigens, the ryanodine receptor, type I interferons and/ or interleukin-12), but not in EOMG (3) (8); iv) inflammation of the thymus, hyperplasia, is common in EOMG. Such thymic hyperplasia is characterised by variable ectopic germinal centre like infiltrates with ongoing AChR autoantibody production, and surgical removal by thymectomy is clinically beneficial. (8) (9) In contrast, since the thymus is normal-for-age, i.e., atrophied, in most cases of LOMG, (10) its removal may be unnecessary. Because the biological age span of EOMG and LOMG can be expected to overlap, biological markers to distinguish the two forms potentially could serve to

identify those that stand to benefit from for example thymectomy, i.e., serving as an intermediate marker, or endophenotype, of the two subforms. However, prior studies exploring restricted immune cell populations, such as follicular helper T and memory T cells subsets, have thus far not identified discriminatory endophenotypes.(11-16) The objective of this study was to explore whether immune cell populations differed in non-thymomatous AChR+ MG of an age span overlapping between EOMG and LOMG. If so, this would strengthen the notion of relevant differences in the immunopathogenesis of EOMG and LOMG, as well as potentially help in channeling patients to the correct treatments.

Results

Study cohorts

To ensure the robustness of our findings, we studied two independent cohorts of EOMG and LOMG patients (Table 1). The UK discovery cohort consisted of 28 untreated AChR+ MG patients, 12 with thymic hyperplasia-associated EOMG (sampled prior to thymectomy but categorised and selected to the study after the procedure) and 16 with LOMG. Samples from patients with ages close to 50 years were preferentially selected and attempts were made to balance the sex ratio, to diminish the bias of the in-built age and sex differences between the patient subgroups on the results. In addition, matched thymic samples were available for 10 patients (9 EOMG and 1 LOMG). Age-matched samples from 20 healthy controls were included (Table 1). Deep spectral cytometry was performed, delineating B cells/CD4 T cells/CD8 T cells/$\gamma\delta$ T cells, and natural killer (NK) cells/innate lymphoid cells, respectively (Supplementary table 1).

The SE validation cohort consisted of 8 EOMG and 6 LOMG patients collected between 2014 and 2020. All had symptomatic generalised MG at sampling with no ongoing immunomodulation,

except one LOMG patient who had received a course of intravenous immunoglobulins four weeks before sampling (Table 1). Five SE EOMG patients had been thymectomized 2-20 years prior, but experienced MG worsening at time of sampling. Samples underwent a single-cell multiomic analysis, including CITE-seq of 9 surface markers, whole transcriptome, and B- and T-cell receptor sequencing.

In addition, samples from five patients (three UK and two SE), considered of uncertain MG subtype, were analysed exploratively. These patients were either younger than 50 years of age but lacked thymic hyperplasia upon pathology examination, or 50 years of age and non-thymectomized. Furthermore, matched pre- and post-thymectomy PBMCs were available for two patients These are shown in figures but not included in statistical analyses.

Main cell subsets and supervised identification of discriminating populations

In order to find immunocellular endophenotypes discriminating the MG subsets, we first investigated differences in major cell populations in the two cohorts by a combination of conventional and Euclidean neighbour-enhanced gating (UK cohort Supplementary figures 1,2 and 3; SE cohort Supplementary figure 4). None of the major populations differed significantly, neither when comparing patient groups, nor to their age-matched controls (Figure 1A-B). Next, the main cell populations were gated into subpopulations in the two cohorts (Supplementary figure 1), revealing a differential abundance of some populations in the UK cohort (Supplementary figure 5A-B). After validation in the SE cohort, the remaining findings were lower percentages of naïve as well as CD161+ CD8+ T cells in LOMG compared to EOMG and age-matched controls (Supplementary figure 5A, C). However, the capacity to discriminate EOMG and LOMG at the level of individual patients was unsatisfactory (Supplementary figure 5D-E). To increase the discriminative capacity, we therefore designed and employed a supervised, differential abundance-

directed cell-selection method, related to previously published methods, (17) but generalised to three groups. In total, 5% of ILC, 8% of CD4 T cells and B cells, 10% of NK cells, 20% of CD8 T cells and 29% of $\gamma\delta$ T cells from the UK cohort data were identified as potentially differentially abundant between EOMG, LOMG and/or control groups. These were further clustered into 61 sub-clusters (exemplified in Figure 1C). Through neighbour-based cluster label transfer in a Euclidean space of common markers for the two cohorts, 11 of these candidate clusters were also identified in the SE cohort data in sufficient numbers (>50 cells) to allow for statistical inference (exemplified in Figure 1D). Next, the frequency of these 11 remaining clusters was compared between EOMG and LOMG, and between EOMG or LOMG and controls in the UK cohort, as well as between EOMG and LOMG in the SE cohort. A cluster was selected if significantly different in all three comparisons (Exemplified in Figure 1E).

Identification of two CD8 T cell and one NK cell cluster as robustly differentially abundant

As expected for any method prone to over-fitting, most clusters identified with the abundance-supervised method were significant only in one or two comparisons. However, three clusters (CD8T#33, CD8T#34 and NK#49) differed in all three comparisons (Figure 2A-B, Table 2). In the UK/SE cohorts respectively, these clusters comprised 9%/16% (CD8T#33) and 3%/2% (CD8T#34) of all CD8T and 6%/8% (NK#49) of total NK cells (Table 2). In line with the results of the initial manual gating analysis (Supplementary figure 5), no distinct B or CD4T cell populations differed, while the frequency of two CD8 T cell populations discriminated between MG subsets and controls. Both CD8 T clusters were smaller in LOMG, i.e., $LOMG^{low}$, as compared to EOMG and controls. These two CD8 T populations could be distinguished with a few

markers; the naïve CD8T#33 cluster by CCR7 positivity alone, and CD8T#34 by being $CD45RA^{low}CD7^{low}CD161^{+}$ (Figure 2C, Table 2 and Supplementary figure 6).

In addition, a cluster of NK cells, NK#49, was reduced in frequency in EOMG compared to LOMG and age-matched controls, i.e., $EOMG^{low}$. This cluster displayed a canonical, highly differentiated NK cell phenotype and could be distinguished from other NK cells by a combination of CD16, CD57, NKp30, CD2 and NKG2C (Figure 2C, Table 2 and Supplementary figure 6). Thus, based on a three-way comparison in the UK cohort replicated in the SE cohorts, we uncovered three differentially abundant cell populations.

Whole transcriptomic and T-cell receptor profiling

The SE cohort was used for in-depth transcriptomic characterization of the three identified cell populations. First, expression of the three top genes in each population validated a corresponding transcriptomic signal for the imputed surface proteins (Figure 3A). SingleR was applied to annotate known cell types, (18) indicating that CD8T#33 comprised naïve $CD8^{+}$ T cells (93%), CD8T#34 mucosa-associated invariant T (MAIT)-cells (87%), and NK#49 conventional $CD56^{dim}$ NK cells (91%) (Figure 3B). To further characterise the latter, inhibitory Killer Immunoglobulin-like Receptor (KIR) gene expression patterns were selectively analysed, showing expression of KIR2DL1, -3DL1, -3DL2 and -2DL3, and low KIR2DL4 and -3DL3 expression, indicating a highly differentiated status (Supplementary figure 7).

Next, compared to their corresponding cell types, differentially regulated genes for each of the clusters were identified (up/down); (19) 401/370 for CD8T#33, 31/72 for CD8T#34, and 13/25 for NK#49 (Figure 3C, Supplementary table 2). In a gene ontology analysis using EnrichR, (20)

cluster CD8T#33 was associated with T cell subset defects, CD8T#34 with broader hematopoietic defects and NK#49 with B cell differentiation and decidual NK cells (Figure 3D, Table 2). (21) As expected, T cell receptor (TCR) characteristics of the naïve CD8T#33 cluster revealed a lower rate of clonality compared to the whole CD8+ T cell population (adjusted Fisher´s exact test $p<0.00001$), while CD8T#34 did not (adjusted Fisher´s exact test $p=1$), but displayed an increase of TCR-Vβ6 and reduction of TCR-Vβ7 (adjusted Fisher´s exact test $p<0.00001$ in each case; Figure 3E, Table 2). Further, CD8T#34 TCR-α gene usage confirmed enrichment of MAIT cells (TCR-Vα7.2 and TCR-Jα33 in 26% of cells vs 0.4% in the entire CD8+ T cell population; adjusted Fishers exact test $p<0.00001$, Figure 3F, Table 2), with the most common prominent TCR-α phenotype being non-clonal MAIT cells expressing TCR-Vβ6 (11% of cells in CD8T#34 vs 0.005% in the entire CD8+ T cell population; Fisher´s exact test $p<0.00001$, Figure 3F, Table 2). In contrast, when comparing EOMG and LOMG in the SE cohort, the overall B cell receptor heavy and TCR-beta gene V-family usage and clonality, and B cell receptor heavy gene mutation rates did not differ noticeably (data not shown).

Relation to thymic pathology and capacity to discriminate MG subgroup membership

Given the role of thymic hyperplasia in EOMG, we next investigated relationships between blood and thymic lymphocyte NK cells and the NK#49 cluster and thymic hyperplasia in a sub-cohort of 10 patients with paired samples available. There was a significant correlation between the NK#49 populations in the two compartments (Spearmans´s rho 0.83, $p=0.006$, Fig. 4A, Table 2), which however was not present for the total NK cell population (Spearmans´s rho 0.33, p-value 0.34, Fig. 4B). There was also a trend towards a negative correlation between grade of hyperplasia and the logarithmic relative size of NK#49 in blood and thymic cell pools (Pearson´s correlation coefficients -0.59 and -0.66 respectively, adjusted p-values 0.1 for both, Fig 4C, Table 2), which

was considerably less pronounced for the entire NK cell population (Pearson´s correlation coefficients -0.42 and -0.41, adjusted p-values 0.32 and 0.32, for logarithmic percentages of blood and thymic cells, respectively, Fig 4C). These findings point to a potential protective role for NK#49 in the control of thymic hyperplasia. Interstingly, the LOMG group showed a wide distribution of this subset, with partial overlap with the EOMG group.

Finally, we explored the capacity of the per-donor frequency of the differentially abundant clusters to correctly assign MG subgroup membership, using the UK cohort for training and the SE cohort for validation. Using partial least squares discriminant analysis, the optimally EOMG-LOMG separating vector based on the frequency of the three clusters was identified. This vector was influenced 35% by the frequency of cluster CD8T#33, 56% by CD8T#34 and 9% by NK#49 (Table 2 and Figure 4G). When it was used for classification of the UK cohort, 10/12 EOMG and 16/16 LOMG could be correctly classified (specificity of 100% and sensitivity of 83% for EOMG) (Figure 4D-E). When the same vector and threshold was used for validation in the SE cohort, 8/8 EOMG and 5/6 LOMG samples were correctly classified (specificity 83%, sensitivity 100% for EOMG) (Figure 4F). Predicting if a control individual was younger or older than 50 using the same vector and threshold was less successful (sensitivity 100% but specificity only 60% for young controls), indicating that the variance in the identified population frequencies is not only explained by age differences. Notably, 6/7 of the post-thymectomy EOMG samples, and 2/3 and 2/2 of the uncertain cases in the UK and SE cohorts, respectively, were categorised as EOMG with this method. The single UK post-thymectomy EOMG sample that was wrongly categorised was slightly above the threshold and had moved 2% of the PLS-DA range from its pre-thymectomy value (Figure 4D). In summary, the identified cell populations (summarised in

figure 4G) differentiate EOMG and LOMG with reasonable sensitivity and specificity without consideration of any other discriminating factors.

Discussion

Leveraging two independent MG cohorts, spanning the traditional EOMG and LOMG age-based delineation cut-off at 50 years of age, we explored the potential existence of immune cell profiles discriminating between these two major MG subgroups. Somewhat contrary to our expectations, major immune cell populations revealed only minor differences, where more sensitive supervised clustering only identified three divergent cell populations. This may partly reflect our conservative study design; a less conservative approach likely would have resulted in a larger number of divergent cell populations, while also increasing the risk of spurious findings. Nevertheless, the identified diverging cell clusters substantiate the notion of differences in underlying immune pathogenesis between EOMG and LOMG.

Under-representation of naïve $CD8^+$ T cells in LOMG aligns with immune senescence, known to increase risk of autoimmunity. (22) (23) Notably, lower frequencies of MAIT cells in blood is linked to aging, (24) and is also a feature associated with several autoimmune disorders. (25) Thus, these cellular profiles of LOMG may reflect exaggerated immune ageing compared to age-matched controls. A previous EOMG study has found blood $CD161^+$ T cells, which includes the MAIT population, to be lower in EOMG compared to age-matched controls, (16) but this observation could not be replicated here.

The size of the EOMG-distinguishing NK#49 cell cluster was strongly correlated in blood and thymic cell populations and showed a negative correlation trend in relation to thymic hyperplasia grades. NK cells have been proposed to play a regulatory role in autoimmune responses, (26) and evidence of $CD56^{dim}$ NK cell dysregulation in autoantibody-mediated neurological disorders,

including MG, has been noted previously. (27) Further, a recent study of expression quantitative trait locus mapping found that several instances of regulatory variation specific to NK cells were associated with risk of autoimmunity. (28) It may therefore be speculated if there is a more complex background to the unusually strong HLA-B*08:01 linkage than involving CD8+ T cell functions, (29) especially since the HLA-B*08 haplotype has been shown to be associated with decreased NK cell functions, and that certain KIR interact with HLA class I molecules.(30, 31) Thymic NK cell frequencies in MG were captured in two previous studies, (15, 32) but as subsets of NK cells were not described in any of the studies, results cannot be directly compared. This also holds for studies describing generally attenuated $CD56^{dim}$ NK cells, with relative increases of $CXCR5^{+}$ NK cells, and decreases in NK cell cytotoxicity. (33) (34) Of note, this population showed a wide frequency distribution among LOMG. Unfortunately, due to lack of comparable disease activity scores from sampled individuals, we were not able to explore if the size of the NK#49 cluster, either in EOMG or in LOMG patients, was associated also with disease activity, as has been shown for a non-canonical NK cell subset in a recent study.(27) Other recent studies have investigated the blood and thymus lymphocyte compartments in MG, finding altered activated $CD4^{+}$ T cell cytokine responses, (15) increases of group 2 innate lymphoid cells and $CD27^{-}$ TCR-γδ T cells, (16) as well as changes in follicular helper T-cells (Tfh). (12) No attempt to validate these findings were done here as no stimulation protocol was included and Tfh frequencies only indirectly could be inferred based on expression in the SE dataset.

Pending validation by future studies, the identified cellular immune endophenotypes may have potential clinical relevance. For example, robust evidence for the efficacy of thymectomy for non-thymomatous MG derives from a single randomized controlled trial showing benefit for

patients with MG aged 18-65 compared to prednisolone alone on 3-year disease severity outcomes, (9) and with higher probability of minimal disease manifestations at 5-years. (35) However, due to the composition of the study population, a beneficial effect in men, individuals aged ≥50 and those with mild disease remain unknown. Further, a retrospective Japanese study showed that only a minority of MG patients aged ≥50 undergoing thymectomy displayed hyperplasia. (10) Still, these displayed less active disease at follow up compared to without hyperplasia, which lends support to the notion that thymectomy is effective with thymic hyperplasia regardless of age. In the present study, a score based solely on the discriminatory cluster frequencies is presented, which separates patients with thymic hyperplasia with a 90% accuracy. Interestingly, one patient from each cohort, both being exactly 50 years of age at the time of diagnosis, not offered thymectomy and therefore not classified as either EOMG or LOMG, but uncertain MG subtype, clustered with the EOMG patients. Even if this does not prove they would have benefited from thymectomy, it underscores that attempts to use case stratification based on biological features for channelling to thymectomy are warranted.

Importantly, if a stratification score would be constructed in a clinical setting, this immunophenotypic information would be integrated with other known factors known to differ between EOMG and LOMG, e.g., levels of AChR antibody titres, sex and, obviously, age. In addition, the three identified cell clusters shed light on underlying disease processes that serve to be further explored in future studies. Our findings may also prove useful for characterizing immune phenotypes and tracking therapeutic responses, not least with potential future therapies relying on tolerogenic approaches. (36)

There were several limitations to our study. First, cohorts were small, affecting our ability to identify discriminatory clusters and potentially also impacting external validity. In light of this,

three precautions were taken. First, a high-resolution, supervised clustering method was employed, sensitively identifying cell clusters with endophenotypic discriminatory potential. Second, to limit over-fitting, nearest neighbors of UK candidate cluster cells were sought in the SE cohort. In this way, 11 of the 61 clusters were identified in both cohorts. Third, a nested statistical test approach followed, identifying three of the 11 clusters as robustly discriminatory. Notably, when ranking the 61 clusters by size, the 11 clusters were interspersed among the largest 27 of the UK clusters (Supplementary table 3). This underscores that a majority of the UK clusters were likely modeled from noise, but also indicates another study limitation, namely the relatively low number of collected cells per individual (n=4,500) in the SE cohort. This likely explains why none of the 34 smaller UK clusters were reproduced in the SE cohort. Additionally, among the three discriminatory clusters, only 146 cells were present in the SE cohort MAIT cluster. These few cells were well-distributed over the EOMG donors, but this finding needs further external validation. On the other hand, in terms of robustness, cell populations that can be detected even with limited cell numbers have a higher utility in a clinical setting. Other limitations includes that many EOMG samples were collected years before the controls in the UK cohort, although no discernible differences in cell viability relating to storage time could be seen. Further, 5 of 8 EOMG in the SE cohort were collected post-thymectomy. However, all displayed relevant clinical worsening of generalized MG symptoms that prompted *de novo* institution of immunotherapy. Still, populations differing purely during the pre-thymectomy phase would have been excluded with our nested statistical approach. The lack of TCR sequencing for the UK cohort also meant that previously identified differences in TCRVβ chain family usage could not be reproduced,(37) in the absence of a correlated surface phenotype. Finally, thymus pathology status was not available for most LOMG individuals. This might have

lowered the magnitude of difference between the groups, as some older individuals with thymic hyperplasia could have been present in the LOMG group. Finally, our work is wholly descriptive in nature, and further validation and mechanistic work is needed to address the discriminatory potential and usefullness of the identified cell subsets.

In conclusion, we identified a highly differentiated NK cell population being diminished in EOMG, negatively correlating with degree of hyperplasia, as well as one naïve CD8+ T and one MAIT cell population that may reflect immune senescence in LOMG. If reproduced by others, these findings highlight the potential of immunological endophenotypes to rationalize treatment approaches, likely also relevant for future therapeutic strategies, such as efforts to develop tolerogenic therapies to suppress reactivity towards AChR epitopes.

Methods

Sex as a biological variable

As the sex ratio is 2:1 in EOMG and 1:1 in LOMG, both male and female patients have been studied in this context. Two methods are used to circumvent that the identified cell populations are associated to sex differences rather than MG subgroup. First, when selecting the cases for the UK cohort, a less sex-biased set of samples has been chosen. Secondly, the healthy controls, especially for EOMG, have been chosen with a more extreme sex bias than the cases (the controls are therefore not said to be matched for sex). This means that the UK EOMG samples will be compared with the somewhat male-dominated LOMG group as well as the female-dominated young healthy controls. With this setup, a difference between the groups is unlikely to arise from sex alone.

Study participants

The United Kingdom (UK) discovery cohort consisted of 12 EOMG, 16, LOMG, 3 patients with uncertain MG subgroup membership (see results section for definition of this last group) and 20 age matched controls. Of these, 8 EOMG, 1 LOMG and 2 uncertain MG subgroup patients and all controls were recruited in Oxford. The Oxford patients were recruited between 1982 and 2002, whereas the controls were recruited between 2019 and 2023. The remaining 4 EOMG, 15 LOMG and one uncertain patient were recruited in Nottingham between 2014 and 2019

The Swedish validation cohort (SE) comprised 8 EOMG, 6 LOMG and 2 patients with uncertain MG subtype recruited between 2014 and 2020.

Procedures for cell freezing, batching, cytometry and sorting

Peripheral blood mononuclear cells (PBMC) were separated within six hours of sampling and slowly frozen in media with 40-90% foetal bovine serum and 10% dimethylsulfoxide (DMSO). Fresh thymic tissue was dispersed mechanically without enzymes, with cells being washed and frozen at 5°C per min in 95% foetal bovine serum and 5% DMSO. This was done by Professor Willcox and coworkers in Oxford, in immediate conjunction with thymectomies.

For the UK cohort, all thawed samples were stained on the same day, to avoid batch effects. The samples were thawed stepwise in successively warmer DNAse-containing medium in batches containing samples of all groups to introduce technical artefacts. All stainings were performed in a staged manner, to minimise the time differences in each staining step for the samples. A 5-step staining protocol was used for both panels, the details of which can be found at https://github.com/jtheorell/EO_vs_LOMG/tree/main/Data/Oxford/Documentation. Markers for both panels can be found in Supplementary table 1. The data were manually acquired on a five

laser Cytek Aurora (Fremont, California) instrument and the files saved raw. Despite 40-year time frame from the first to the last collected sample, no impact of storage time on the quality of cells could be discerned (Supplementary figure 8).

For the SE cohort, 8 samples were stained and processed on each of two consecutive days. On each day, samples were thawed slowly in warm medium in a mixed order, so that every second sample was an EOMG sample. This was followed by staining with cite-seq (CITE) antibodies (T/B/NK TotalSeq panel from BioLegend, San Diego, California), including hashing antibodies specific to each sample, as well as with DAPI for dead cell detection. Markers for cite-seq panel is given in Supplementary table 4. Sorting was conducted on a BD Influx System (Franklin Lakes, New Jersey) instrument, selecting live singlet mononuclear leukocytes (Supplementary figure 9). The whole sorting procedure took 38 and 31 minutes for the two experimental days, respectively. Details of the protocols and further sort documentation for the two experimental days can be found at https://github.com/jtheorell/EO_vs_LOMG/tree/main/Data/Stockholm/Sort_documentation.

Transcriptome, adaptive immune receptor repertoire (AIRR) and surface marker (cite-seq) sequencing

On each of the two experimental days, live sorted PBMC were immediately brought to the Eukaryotic Single Cell Genomics Facility at SciLife Lab, Stockholm. Here, the Chromium 5' version 3 kit (10X Genomics, Pleasanton, California) was used, in combination with protocols for T- and B cell receptor sequencing, as well as 9-marker cite-seq, including 8-sample hashing. This allowed for pooling of all samples from each experimental day, minimising technical artefacts. Quality controls and further documentation about the processing can be found at

https://github.com/jtheorell/EO_vs_LOMG/tree/main/Data/Stockholm/10X_quality_controls_and_documentation.

Handling of raw spectral flow cytometry standard (FCS) files and gating for the UK cohort

After export of the raw FCS files from the Cytek Aurora instrument, files were unmixed using the flowSpecs package in R, (38) including a manual curation step of the parameter names. After this, the files were imported into FlowJo (BD) and the unmixing artifacts were corrected with a manual correction step, starting out with a square 0 matrix, which was filled with small corrective values upon visual identification of populations with spectral unmixing-related spurious correlations (similar to over- or under-compensation in conventional cytometry). After this step, gating was conducted in FlowJo in two ways; first, with a full, conventional gating strategy as outlined in Supplementary figures 1-3, and subsequently with a few gates, excluding debris, dead cells and identifying the main cell populations (gates in Supplementary figure 1, 2 and 4). These initial FlowJo-based analyses was followed by bioinformatic and statistical analyses in R, including further identification of the identified discriminant clusters through threshold filtering, akin to conventional gating.

Bioinformatic analyses

The full bioinformatic workflow can be accessed at https://github.com/jtheorell/EO_vs_LOMG/tree/main/Scripts, with a narrative description of the analyses included in the supplementary methods section.

Statistics

For all analyses concerning differences between groups, Wilcoxon rank-sum tests are used. Samples an uncertain MG subtype, thymus populations and UK post-thymectomy PBMC samples were not included unless explicitly stated. Notably, for the SE cohort, PBMC obtained also post-thymectomy were included in the statistical analyses. Use of other statistical test have been specified in the text, e.g., Pearson correlation, Spearman´s rank correlation coefficient and Fisher´s exact test. The reason for the use of Pearson correlation for tests involving the ordinal thymic hyperplasia scores is that rank correlations perform poorly when multiple values are identical.

For the nested, supervised clustering method aimed at identifying discriminating clusters, statistical difference was defined by a one-tailed p-value <0.05 for all three comparisons (UK EOMG vs LOMG and EOMG/LOMG vs control, and SE EOMG vs LOMG. The direction of the tail was defined by the population size in the patient group of interest compared to the other patient group and the controls. In other words, a population selected due to its smaller size in EOMG was only statistically tested to be smaller in EOMG compared to the other groups in both cohorts. This setup with three independent tests lowered the false positive error rate to 0.05^3 = 0.000125 or one in 8,000 tests. If we had not used this nested approach, only making 11 comparisons in total, and would have used the strictest correction for multiple comparisons, Bonferroni correction, on a p-value threshold of 0.05, the corrected p-value would have been 0.05/11 = 0.0045 or one false positive in 220 tests, which was considered too high a risk with the dataset sizes in question. Had we instead combined the nested approach with further testing for multiple comparisons, the result would have been 0.000125/11 or one false positive in 88,000 tests, which on the contrary would be too conservative. Thus, our current strategy contains a

strict p-value selection threshold and no additional adjustment for multiple comparisons was made for the nested identification of discriminatory clusters. A false detection rate procedure was applied for all other multiple comparisons.

Study approval

All individuals recruited in Oxford gave oral consent (keeping only information about their age and sex); as approved by the Central Oxford Research Ethics Committee (COREC 1702). The Nottingham patients provided written consent, as approved by the Yorkshire and Humber Research Ethics Committee (REC16/YH/0013). The Swedish patients provided written consent, as approved by the Swedish Ethical Review Authority (dnr 2009/2107-31/2, last amended 2024-07377-02, and dnr 2016/827-31, last amended 2024-02236-02).

Data Availability

A file with the data for all subfigures, called "Supplementary_file_5_all_figure_data.xlsx" is available. If a reader would like to reproduce the full analysis worflow and get all the pre-processed data, all the underlying data for the whole analysis workflow, including the raw count matrices as well as all data resulting from the initial spectral cytometry analyses can be accessed at https://github.com/jtheorell/EO_vs_LOMG/tree/main/Data/Stockholm/Raw_data, where the SE dataset is split into 3 files, due to size restrictions. Available here is also the Kotliarov data set. (39)

All unmixed UK cohort FCS files can be directly downloaded from http://flowrepository.org/id/FR-FCM-Z86X. Raw fastq files from the SE cohort can be viewed at https://doi.org/10.48723/9gg8-ry45. Only researchers who have an ethical permit can apply for access to these files, given European personal data protection laws.

Acknowledgements

The authors acknowledge support from the National Genomics Infrastructure in Stockholm, funded by Science for Life Laboratory and the Knut and Alice Wallenberg Foundation, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Specific funding was received from the EJP-RD Joint Translational Program 2023 (OptiMyG, grant no. 2023-00533), the Swedish Brain Fund, the Region Stockholm (grant no. FoUI-987565), the Swedish Research Council (grant no. 2023-00533), a senior clinical fellowship from the Kogod Centre on Aging, Medical Research Council [MR/V007173/1], Wellcome Trust Fellowship [104079/Z/14/Z], and by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). We acknowledge the work by late Professor J Newsom-Davis and Sister E Goodger, University of Oxford, Oxford, UK, with samples from the 1986-2002 UK patients and Prof N Willcox, University of Oxford, Oxford, UK, for sample curation and suggestions for the project and manuscript. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Author Contributions (according to CRediT)

JT: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration; **NR**: Validation, Investigation, Writing - Review & Editing; **AF**: Resources, Writing - Review & Editing; **CS**: Formal analysis, Writing - Review & Editing, Visualization; **PA**: Resources, Writing

- Review & Editing; **VD**: Writing - Review & Editing; **LH**: Resources, Writing - Review & Editing; **IL**: Resources, Writing - Review & Editing; **SRI**: Validation, Resources, Writing - Review & Editing, Supervision; **SB**: Validation, Resources, Data Curation, Writing - Review & Editing; **AEH**: Validation, Resources, Data Curation, Writing - Review & Editing; **FP**: Conceptualization, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition.

References

1. Meriggioli MN, and Sanders DB. Autoimmune myasthenia gravis: emerging clinical and biological heterogeneity. *Lancet Neurol.* 2009;8(5):475-90.
2. Gilhus NE. Myasthenia Gravis. *N Engl J Med.* 2016;375(26):2570-81.
3. Gilhus NE, Skeie GO, Romi F, Lazaridis K, Zisimopoulou P, and Tzartos S. Myasthenia gravis - autoantibody characteristics and their implications for therapy. *Nat Rev Neurol.* 2016;12(5):259-68.
4. Gilhus NE, Tzartos S, Evoli A, Palace J, Burns TM, and Verschuuren J. Myasthenia gravis. *Nat Rev Dis Primers.* 2019;5(1):30.
5. Gilhus NE, and Verschuuren JJ. Myasthenia gravis: subgroup classification and therapeutic strategies. *Lancet Neurol.* 2015;14(10):1023-36.
6. Alvarez-Velasco R, Gutierrez-Gutierrez G, Trujillo JC, Martinez E, Segovia S, Arribas-Velasco M, et al. Clinical characteristics and outcomes of thymoma-associated myasthenia gravis. *Eur J Neurol.* 2021;28(6):2083-91.
7. Braun A, Shekhar S, Levey DF, Straub P, Kraft J, Panagiotaropoulou GM, et al. Genome-wide meta-analysis of myasthenia gravis uncovers new loci and provides insights into polygenic prediction. *Nat Commun.* 2024;15(1):9839.
8. Hapnes L, Willcox N, Oftedal BE, Owe JF, Gilhus NE, Meager A, et al. Radioligand-binding assay reveals distinct autoantibody preferences for type I interferons in APS I and myasthenia gravis subgroups. *J Clin Immunol.* 2012;32(2):230-7.
9. Wolfe GI, Kaminski HJ, Aban IB, Minisman G, Kuo HC, Marx A, et al. Randomized Trial of Thymectomy in Myasthenia Gravis. *N Engl J Med.* 2016;375(6):511-22.
10. Uzawa A, Kawaguchi N, Kanai T, Himuro K, Oda F, Yoshida S, et al. Two-year outcome of thymectomy in non-thymomatous late-onset myasthenia gravis. *J Neurol.* 2015;262(4):1019-23.
11. Sagar HJ, Davies-Jones GA, and Allonby ID. Clinical and immunological associations in myasthenia gravis. 2. Cell-mediated immunity. *J Neurol Neurosurg Psychiatry.* 1980;43(11):971-7.
12. Ashida S, Ochi H, Hamatani M, Fujii C, Kimura K, Okada Y, et al. Immune Skew of Circulating Follicular Helper T Cells Associates With Myasthenia Gravis Severity. *Neurol Neuroimmunol Neuroinflamm.* 2021;8(2).

13. Zhang CJ, Gong Y, Zhu W, Qi Y, Yang CS, Fu Y, et al. Augmentation of Circulating Follicular Helper T Cells and Their Impact on Autoreactive B Cells in Myasthenia Gravis. *J Immunol.* 2016;197(7):2610-7.
14. Yang Y, Zhang M, Ye Y, Ma S, Fan L, and Li Z. High frequencies of circulating Tfh-Th17 cells in myasthenia gravis patients. *Neurol Sci.* 2017;38(9):1599-608.
15. Ingelfinger F, Krishnarajah S, Kramer M, Utz SG, Galli E, Lutz M, et al. Single-cell profiling of myasthenia gravis identifies a pathogenic T cell signature. *Acta Neuropathol.* 2021;141(6):901-15.
16. Verdier J, Fayet OM, Hemery E, Truffault F, Pinzon N, Demeret S, et al. Single-cell mass cytometry on peripheral cells in Myasthenia Gravis identifies dysregulation of innate immune cells. *Frontiers in immunology.* 2023;14:1083218.
17. Zhao J, Jaffe A, Li H, Lindenbaum O, Sefik E, Jackson R, et al. Detection of differentially abundant cell subpopulations in scRNA-seq data. *Proc Natl Acad Sci U S A.* 2021;118(22).
18. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* 2019;20(2):163-72.
19. Chen Y, Chen L, Lun ATL, Baldoni PL, and Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv.* 2024.
20. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14:128.
21. Moffett A, and Colucci F. Uterine NK cells: active regulators at the maternal-fetal interface. *J Clin Invest.* 2014;124(5):1872-9.
22. Ray D, and Yung R. Immune senescence, epigenetics and autoimmunity. *Clin Immunol.* 2018;196:59-63.
23. Liu Q, Zheng Y, Goronzy JJ, and Weyand CM. T cell aging as a risk factor for autoimmunity. *J Autoimmun.* 2023;137:102947.
24. Novak J, Dobrovolny J, Novakova L, and Kozak T. The decrease in number and change in phenotype of mucosal-associated invariant T cells in the elderly and differences in men and women of reproductive age. *Scand J Immunol.* 2014;80(4):271-5.
25. Chiba A, Murayama G, and Miyake S. Mucosal-Associated Invariant T Cells in Autoimmune Diseases. *Frontiers in immunology.* 2018;9:1333.
26. Flodstrom-Tullberg M, Bryceson YT, Shi FD, Hoglund P, and Ljunggren HG. Natural killer cells in human autoimmunity. *Curr Opin Immunol.* 2009;21(6):634-40.
27. Yandamuri SS, Filipek B, Lele N, Cohen I, Bennett JL, Nowak RJ, et al. A Noncanonical CD56dimCD16dim/- NK Cell Subset Indicative of Prior Cytotoxic Activity Is Elevated in Patients with Autoantibody-Mediated Neurologic Diseases. *J Immunol.* 2024;212(5):785-800.
28. Gilchrist JJ, Makino S, Naranbhai V, Sharma PK, Koturan S, Tong O, et al. Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease associations, highlighting their role in autoimmunity. *Nat Commun.* 2022;13(1):4073.
29. Saunders PM, Vivian JP, Baschuk N, Beddoe T, Widjaja J, O'Connor GM, et al. The interaction of KIR3DL1*001 with HLA class I molecules is dependent upon

molecular microarchitecture within the Bw4 epitope. *J Immunol.* 2015;194(2):781-9.
30. Caruso C, Candore G, Colucci AT, Cigna D, Modica MA, Tantillo G, et al. Natural killer and lymphokine-activated killer activity in HLA-B8,DR3-positive subjects. *Hum Immunol.* 1993;38(3):226-30.
31. Gregersen PK, Kosoy R, Lee AT, Lamb J, Sussman J, McKee D, et al. Risk for Myasthenia Gravis Maps to a (151)Pro -> Ala Change in TNIP1 and to Human Leukocyte Antigen-B*08. *Annals of Neurology.* 2012;72(6):927-35.
32. Terroba-Navajas P, Lu IN, Quast I, Heming M, Keller CW, Ostendorf L, et al. Single-Cell Transcriptomics Identifies a Prominent Role for the MIF-CD74 Axis in Myasthenia Gravis Thymus. *Neurol Neuroimmunol Neuroinflamm.* 2025;12(3):e200384.
33. Ge MR, Yang CL, Li T, Du T, Zhang P, Li XL, et al. Circulating CXCR5(+) natural killer cells are expanded in patients with myasthenia gravis. *Clin Transl Immunology.* 2023;12(5):e1450.
34. Zhang Q, Han X, Bi Z, Yang M, Lin J, Li Z, et al. Exhausted signature and regulatory network of NK cells in myasthenia gravis. *Frontiers in immunology.* 2024;15:1397916.
35. Lee I, Kuo HC, Aban IB, Cutter GR, McPherson T, Kaminski HJ, et al. Minimal manifestation status and prednisone withdrawal in the MGTX trial. *Neurology.* 2020;95(6):e755-e66.
36. Lazaridis K, Fernandez-Santoscoy M, Baltatzidou V, Andersson JO, Christison R, Grunberg J, et al. A Recombinant Acetylcholine Receptor alpha1 Subunit Extracellular Domain Is a Promising New Drug Candidate for Treatment Of Myasthenia Gravis. *Frontiers in immunology.* 2022;13:809106.
37. Tackenberg B, Schlegel K, Happel M, Eienbroker C, Gellert K, Oertel WH, et al. Expanded TCR Vbeta subsets of CD8(+) T-cells in late-onset myasthenia gravis: novel parallels with thymoma patients. *J Neuroimmunol.* 2009;216(1-2):85-91.
38. Theorell J. Bioconductor; 2022.
39. Kotliarov Y, Sparks R, Martins AJ, Mule MP, Lu Y, Goswami M, et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat Med.* 2020;26(4):618-29.

Tables

Cohort		UK						SE			
	Group	EO-MG	LO-MG	Un-certain	Young ctrl	Old ctrl	All	EO-MG	LO-MG	Un-certain	All
	Number of individuals	12	16	3	10	10	51	8	6	2	16
	Female %	50	38	33	80	50	51	100	50	50	75
Age at onset	min	16	54	34	21	55	16	29	54	41	29
	median	35	67	36	36	67	54	36	63	46	46.5
	max	45	90	50	43	73	90	49	77	50	77
Collection year	min	1982	1986	1986	2019	2019	1982	2014	2017	2018	2014
	median	1995	2015	1989	2023	2019	2016	2017	2019	2018	2018
	max	2019	2019	2016	2023	2023	2023	2020	2020	2018	2020
PBMC source	From non-thymec-tomised	12	16	3	10	10	51	3	6	2	11
	Paired pre-post-thymecto my	2	0	0	0	0	2	0	0	0	0
	Only post-thymecto my	0	0	0	0	0	0	5	0	0	5
	Paired thymic cells	9	1	0	0	0	10	0	0	0	0
	Immuno-therapy	0	0	0	0	0	0	0	1 IvIG	1 predni-solone	2
	Analysed cells (x1000)	8475	3280	2059	5193	4819	23825	18	12	5	35

Table 1. Patient and sample characteristics.

			Sub-cluster	CD8T#33	CD8T#34	NK#49
Number of cells			UK	365,185	128,071	395,043
			SE	1,242	146	316
Percent of parent population			UK	9	3	6
			SE	16	2	8
Size of cluster	EOMG		Pre thymectomy	Normal	Normal	Small
			Post thymectomy	Normal	Normal	Small
			LOMG	Small	Small	Normal
	Controls		Young	Normal	Normal	Normal
			Old	Normal	Normal	Normal
			Uncertain MG subtype	Normal	Normal	Normal UK/ Small SE
			Phenotype	Naïve	MAIT	Cytotoxic, conventional
			Defining markers	CD45RA/ CCR7^{high}	CD45RA/ CD7^{low}	CD57/NKp30/ CD16^{high} CD2/NKG2C^{low}
			TCR characteristics	Polyclonal	Vα7.2/Jα33/V6 common Vβ7 uncommon	Not applicable
			Top pathway associations	Defects in activation, proliferation and numbers of T cell subsets	Lower thymocyte number and IgG1 levels, abnormal B cell differentiation	Abnormal B cell differentiation, increases in decidual NK cells
			Size correlation thymus-blood	No	No	Yes
Neg correlation to thymic hyperplasia			PBMC	No	No	Trend
			Thymic cells	No	No	Trend
			Contribution to group separation, %	35	56	9

Table 2. Characteristics of discriminant clusters.

Figures

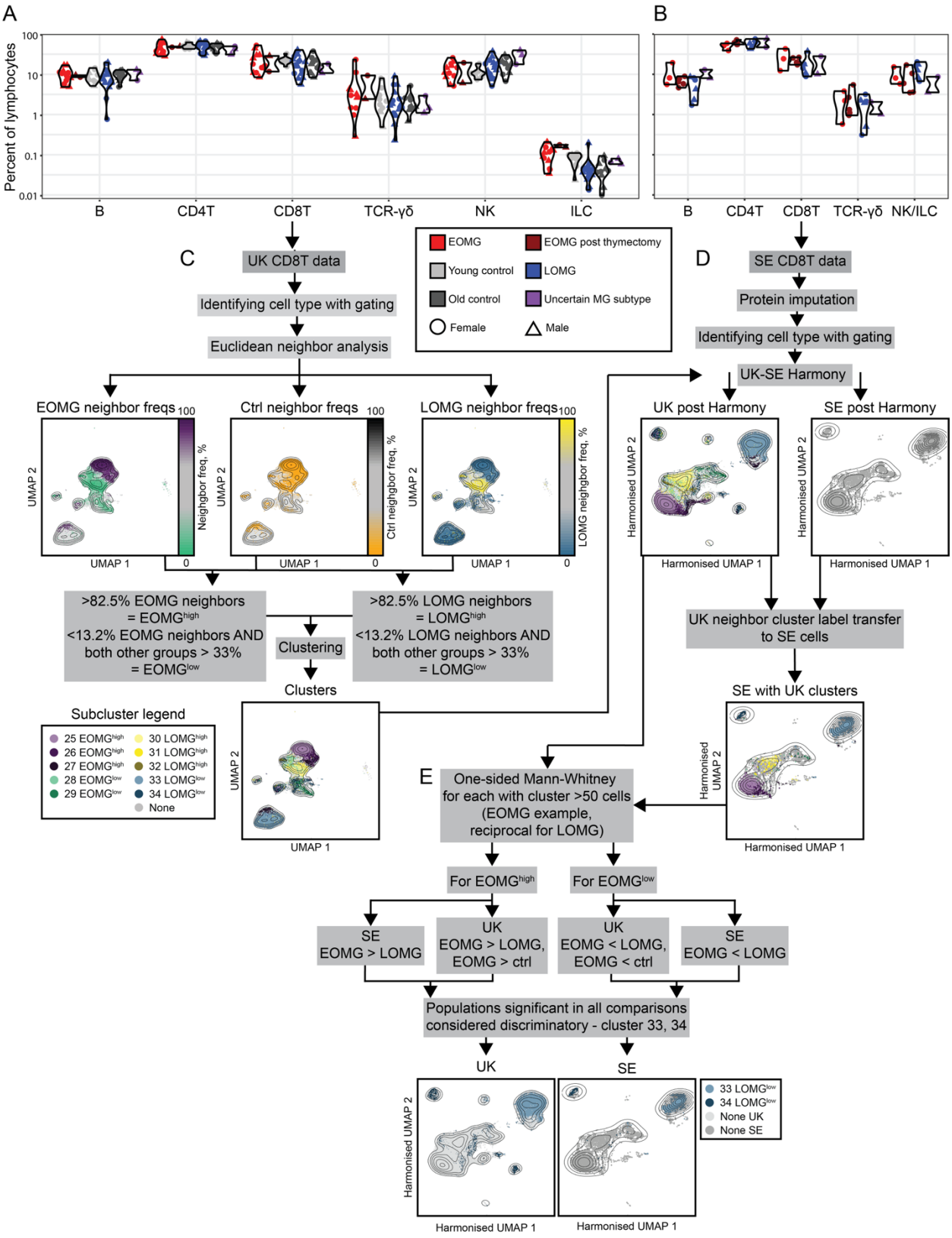
A
B
Percent of lymphocytes
100
10
1
0.1
0.01
B
CD4T
CD8T
TCR-γδ
NK
ILC
NK/ILC
EOMG
EOMG post thymectomy
Young control
LOMG
Old control
Uncertain MG subtype
Female
Male
C
UK CD8T data
Identifying cell type with gating
Euclidean neighbor analysis
D
SE CD8T data
Protein imputation
Identifying cell type with gating
UK-SE Harmony
EOMG neighbor freqs
Ctrl neighbor freqs
LOMG neighbor freqs
Neighbor freq, %
Ctrl neighbor freq, %
LOMG neighbor freq, %
UMAP 1
UMAP 2
UK post Harmony
SE post Harmony
Harmonised UMAP 1
Harmonised UMAP 2
>82.5% EOMG neighbors = EOMG^high <13.2% EOMG neighbors AND both other groups > 33% = EOMG^low
>82.5% LOMG neighbors = LOMG^high <13.2% LOMG neighbors AND both other groups > 33% = LOMG^low
Clustering
Clusters
UK neighbor cluster label transfer to SE cells
SE with UK clusters
Subcluster legend
25 EOMG^high
26 EOMG^high
27 EOMG^high
28 EOMG^low
29 EOMG^low
30 LOMG^high
31 LOMG^high
32 LOMG^high
33 LOMG^low
34 LOMG^low
None
E
One-sided Mann-Whitney for each with cluster >50 cells (EOMG example, reciprocal for LOMG)
For EOMG^high
For EOMG^low
SE EOMG > LOMG
UK EOMG > LOMG, EOMG > ctrl
UK EOMG < LOMG, EOMG < ctrl
SE EOMG < LOMG
Populations significant in all comparisons considered discriminatory - cluster 33, 34
UK
SE
33 LOMG^low
34 LOMG^low
None UK
None SE

Figure 1. Overarching cell type frequencies and overview of supervised clustering workflow. Figure 1A-B: Individual frequencies of the overarching cell types in the UK (A) and SE (B) cohorts. No differences reached significance after adjustment for multiple comparisons. Figure 1C-E: supervised clustering workflow. In this case, CD8 T cells are used as an example, but the workflow is identical for all cell types.

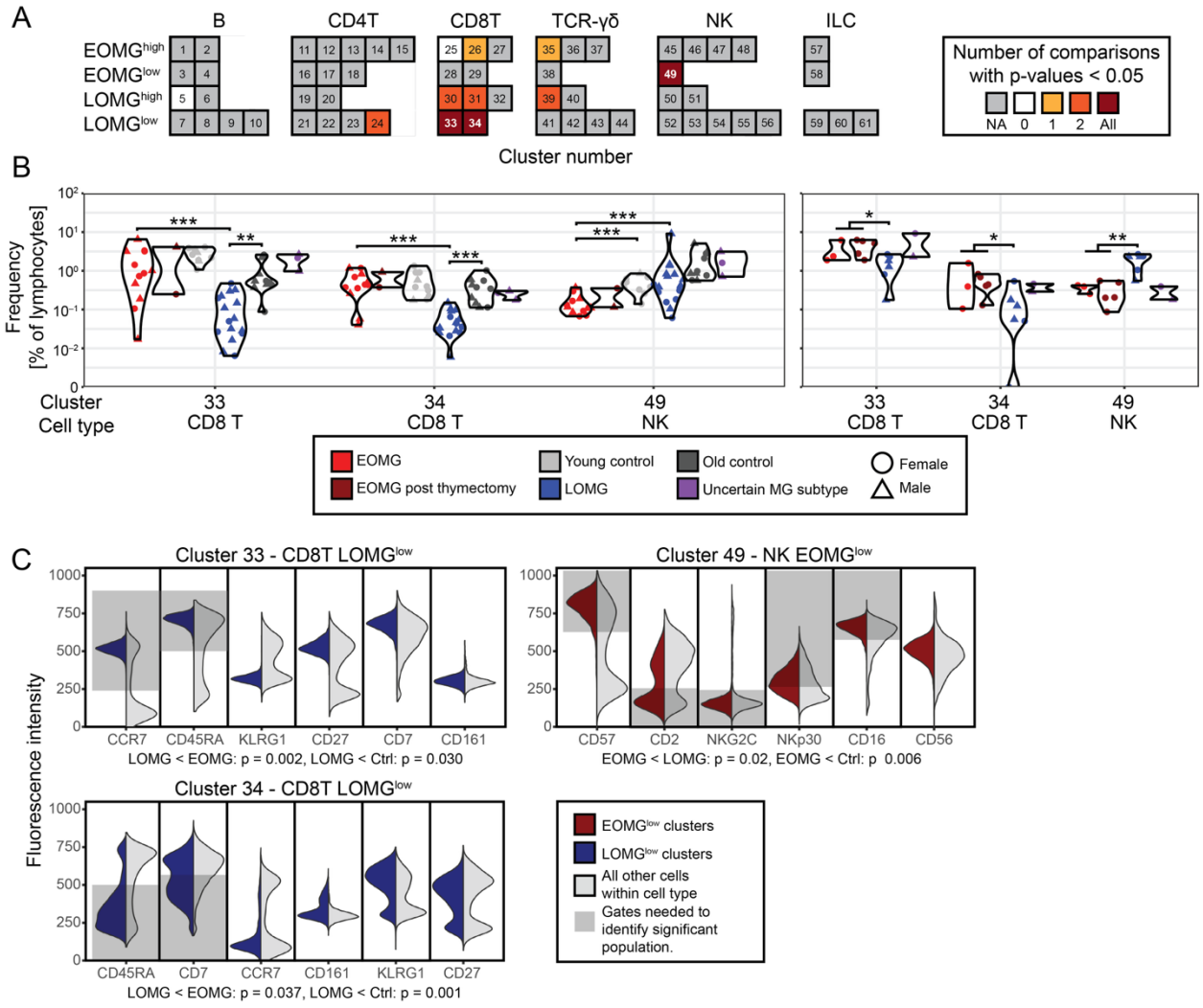

Figure 2. Identification and characterization of significantly different cell populations. A: Overview of analyzed clusters and their respective significance in the iterative process. Dark red indicates significance in all three tests, which was the selection criterion for discriminatory clusters. Grey indicates that <50 cells were present in the SE data, so no comparisons were made. For all clusters where the significance was restricted to one or two populations, these came from the UK cohort. B: individual frequencies of the identified significantly different cell populations in the UK and SE cohorts. Stars indicate level of significance in Mann-Whitney U tests that are one-sided for all comparisons, given that the cell populations were pre-identified as over- or

under-represented in EOMG or LOMG compared to the other groups. Stars above column pairs indicate level of significance with one star corresponding to a p-value of 0.05-0.005, two stars to 0.005-0.0005 and three stars indicating a p-value below 0.0005. As clusters have only been selected if all three independent comparisons had one-tailed p-values lower than 0.05, the false-positive error rate is 1/8000 and with only 11 comparisons in total, no further adjustment for multiple comparisons was therefore included. C: Selected surface markers defining the individual clusters in the UK data set. The grey fields indicate gates that are sufficient to identify the significant cluster.

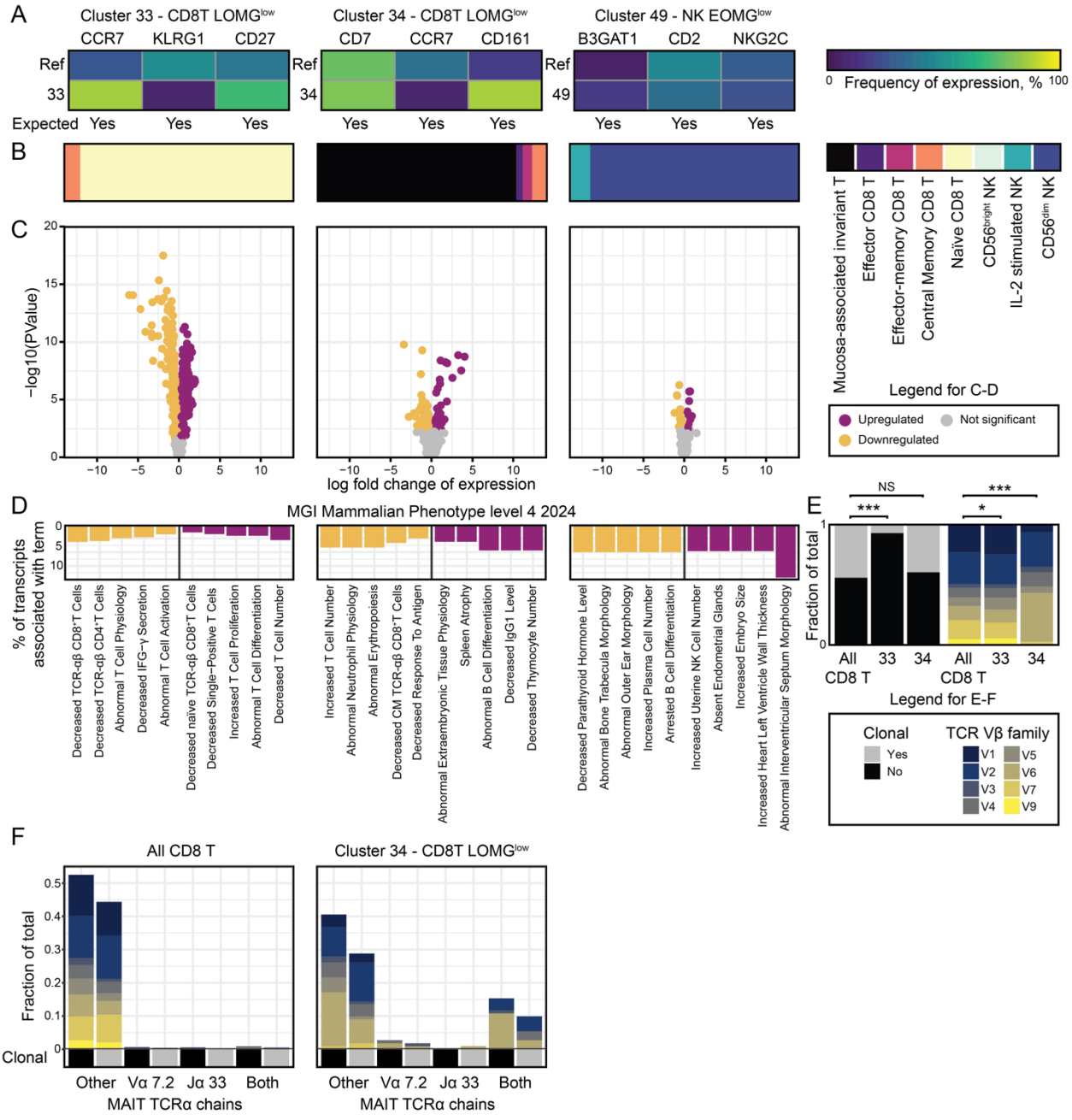

Figure 3. Characterization of significant cell populations. A: Heatmaps indicating fraction of cells expressing genes among the top differentially expressed proteins for each of the three significant populations. Row one shows all cells apart from the population of interest and row two the population of interest. The three top proteins with a corresponding gene were selected for this analysis (the CD45RA isoform of CD45 not being useful in this context). B: SingleR

analysis identifying cluster 33 as naïve CD8 T cells, cluster 34 as MAIT cells and cluster 49 as mature conventional NK cells. C: Volcano plots indicating differentially expressed genes within the three cell subsets. D: Top MGI mammalian phenotype terms associated with the top up- and downregulated genes for the respective cell populations. The length of the column is relative to the fraction of the total number of up- or downregulated genes for the cell subset that are associated to the term. E: Clonality and TCR Vβ chain family usage for the T cell populations. The first subfigure shows the distribution of clonality and TCR Vβ families for all CD8+ T cells, followed by columns for cluster 33 and 34. Stars above column pairs indicate level of significance with one star corresponding to a p-value of 0.05-0.005, two stars to 0.005-0.0005 and three stars indicating a p-value below 0.0005. F: Clonality and TCR Vβ split by MAIT TCRα chain usage. The first subfigure shows the distribution for all CD8+ T cells, whereas the second shows cluster 34. Colour codes are the same as for F.

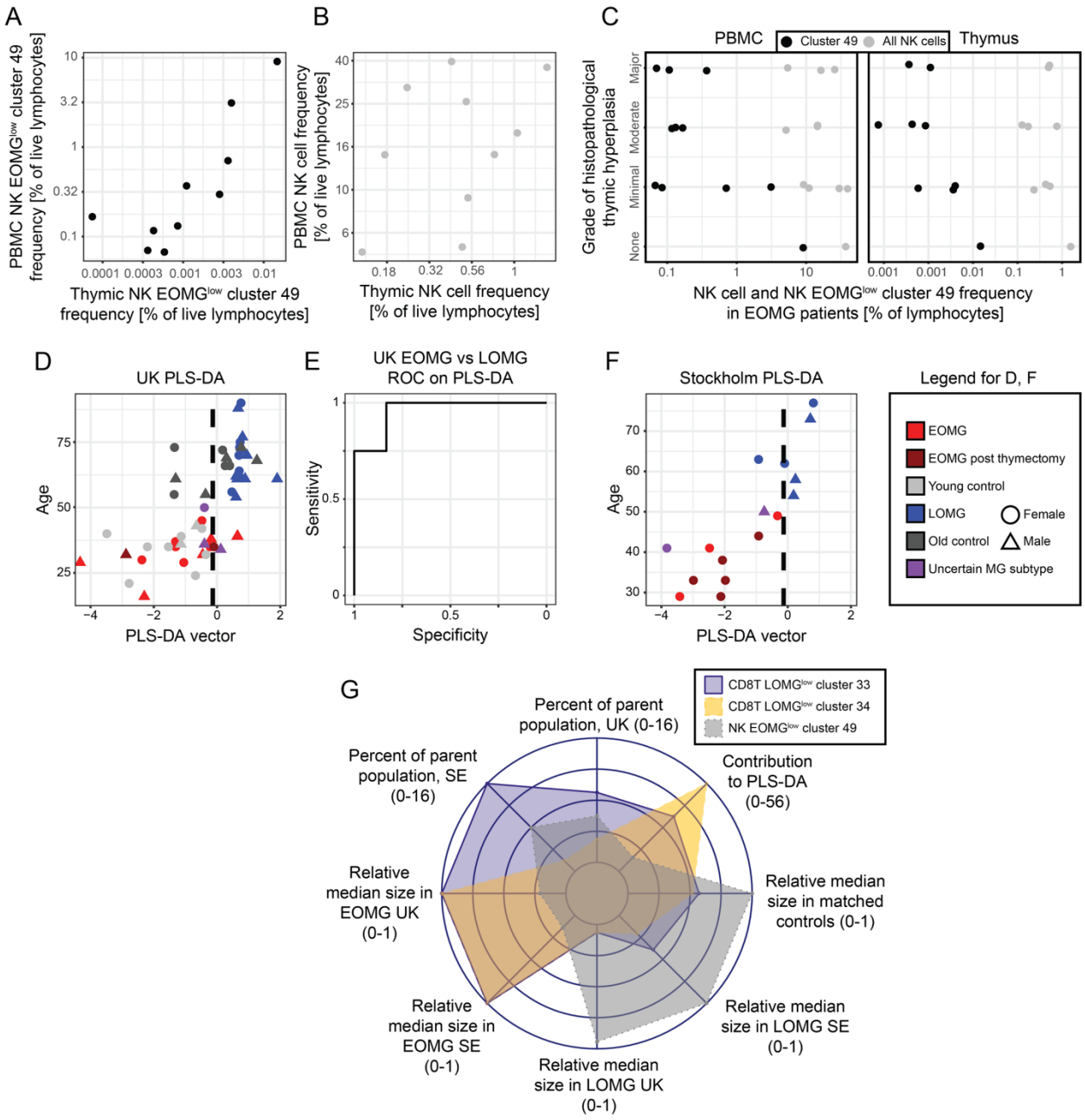

Figure 4. Thymic cell correlations and patient group discrimination. A,B: Frequency of cluster 49 cells (A) and all other NK cells (B) in blood (y-axes) as a function of the same cell population frequencies in thymus. Black dots indicate cluster 49 frequencies, whereas grey dots are the frequencies of NK cells after subtraction of cluster 49. C: Frequency of cluster 49 in PBMC and thymic cells from EOMG patients in the UK cohort correlated to the grade of histopathological

thymic hyperplasia. Colors have the same meaning as in A/B. D: Partial least squares discriminant analysis based on the frequency of the three identified clusters, separating the patient populations in the UK cohort. Model generated only with EOMG and LOMG patients with uncertain myasthenia subpopulation patients, post-thymectomy samples and control groups merely displayed. E: ROC curve based on the PLS-DA for the UK data. F: the SE patient data displayed on the UK PLS-DA model, with the separation threshold inherited from the UKdata. G: Spider plot summing up some of the characteristics of the three identified discriminatory clusters. "Parent population" refers to all CD8 T cells for clusters 33 and 34 and all NK cells for cluster 49.