

RESEARCH

Open Access



# Loss of chromosome Y in regulatory T cells

Jonas Mattisson<sup>1†</sup>, Jonatan Halvardson<sup>1†</sup>, Hanna Davies<sup>1</sup>, Bożena Bruhn-Olszewska<sup>1</sup>, Paweł Olszewski<sup>2</sup>, Marcus Danielsson<sup>1</sup>, Josefin Bjurling<sup>1</sup>, Amanda Lindberg<sup>1</sup>, Ammar Zaghlool<sup>1</sup>, Edyta Rychlicka-Buniowska<sup>2</sup>, Jan P. Dumanski<sup>1,2</sup> and Lars A. Forsberg<sup>1,3\*</sup>

## Abstract

**Background** Mosaic loss of chromosome Y (LOY) in leukocytes is the most prevalent somatic aneuploidy in aging humans. Men with LOY have increased risks of all-cause mortality and the major causes of death, including many forms of cancer. It has been suggested that the association between LOY and disease risk depends on what type of leukocyte is affected with Y loss, with prostate cancer patients showing higher levels of LOY in CD4+T lymphocytes. In previous studies, Y loss has however been observed at relatively low levels in this cell type. This motivated us to investigate whether specific subsets of CD4+T lymphocytes are particularly affected by LOY. Publicly available, T lymphocyte enriched, single-cell RNA sequencing datasets from patients with liver, lung or colorectal cancer were used to study how LOY affects different subtypes of T lymphocyte. To validate the observations from the public data, we also generated a single-cell RNA sequencing dataset comprised of 23 PBMC samples and 32 CD4+T lymphocytes enriched samples.

**Results** Regulatory T cells had significantly more LOY than any other studied T lymphocytes subtype. Furthermore, LOY in regulatory T cells increased the ratio of regulatory T cells compared with other T lymphocyte subtypes, indicating an effect of Y loss on lymphocyte differentiation. This was supported by developmental trajectory analysis of CD4+T lymphocytes culminating in the regulatory T cells cluster most heavily affected by LOY. Finally, we identify dysregulation of 465 genes in regulatory T cells with Y loss, many involved in the immunosuppressive functions and development of regulatory T cells.

**Conclusions** Here, we show that regulatory T cells are particularly affected by Y loss, resulting in an increased fraction of regulatory T cells and dysregulated immune functions. Considering that regulatory T cells plays a critical role in the process of immunosuppression; this enrichment for regulatory T cells with LOY might contribute to the increased risk for cancer observed among men with Y loss in leukocytes.

**Keywords** Regulatory T cells, Tregs, Mosaic loss of chromosome Y, LOY, Single-cell RNA sequencing, scRNAseq, scRNA-seq, Gene set enrichment analysis, GSEA

<sup>†</sup>Jonas Mattisson and Jonatan Halvardson contributed equally to this work.

\*Correspondence:

<sup>1</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

<sup>2</sup>3P-Medicine Laboratory, Medical University of Gdańsk, Gdańsk, Poland

<sup>3</sup>The Beijer Laboratory, Uppsala University, Uppsala, Sweden



## Background

Loss of chromosome Y from hematological progenitors results in mosaic loss of chromosome Y (LOY) in circulating leukocytes, representing the most prevalent somatic mutation in general populations [1–3]. Affected men have an increased risk for mortality and morbidity [4, 5], including all major causes of death such as cardiovascular disease [5–7], cancer [2, 4, 8], Alzheimer's disease [9]. The link between Y loss in blood and disease in other organs is currently being explored, with evidence emerging for direct causality related to the LOY condition. It has for instance been shown that LOY: (I) affects the distribution of blood cell types [10, 11], (II) leads to dysregulation of almost 500 autosomal genes in a cell type dependent manner [12], (III) reduces the abundance of immunoprotein CD99 on the surface of cells, a protein crucial for regulating the permeability of blood vessels [13], (IV) induce fibrosis of internal organs and organ dysfunction in murine models [7], (V) is linked to an expansion of low-density neutrophils during COVID-19 infection [14]. Furthermore, we have previously shown that the impact of LOY on disease risk depends on which leukocyte subset is affected by Y loss [12]. Specifically patients with Alzheimer's disease had more LOY in NK cells, while patients with prostate cancer displayed higher levels of LOY in granulocytes as well as CD4+ T lymphocytes [12]. This however results in an enigma: Y loss in CD4+ T lymphocytes is associated with increased risk of cancer, yet LOY occurs at relatively low levels in CD4+ T lymphocytes compared with other leukocytes [3, 12, 13].

Regulatory T cells (Tregs) are a highly specialized subtype of CD4+ T lymphocytes, constituting less than 10% of circulating CD4+ cells [15, 16]. Hence, the total population of peripheral blood mononuclear cells (PBMC) contains an even smaller fraction of Tregs, which can be difficult to discern from other CD4+ T lymphocytes in single-cell experiments [15, 16]. Yet, Tregs play a critical role in the regulation of immune functions by performing immunosuppression, a balancing act as not to suppress a necessary immune response [17]. Thus, Treg dysregulation has been linked to processes such as long-term inflammation, tissue damage and increased fibrosis [18]. Tregs also contributes to tumour development by directly inhibiting the immune response of surrounding effector cells [19, 20]. Considering the major impact of this rare CD4+ subset, we sought to investigate how the heterogeneity of CD4+ T lymphocytes is affected by LOY. Especially, since it was recently shown that tumours with LOY can influence the distribution, expression and function of T lymphocytes, inducing an immunosuppressive microenvironment that support tumour growth [21]. One of these effects was a higher portion of Tregs in the microenvironment of Y loss bladder cancers [21]. Here, we leverage the power of single-cell RNA sequencing

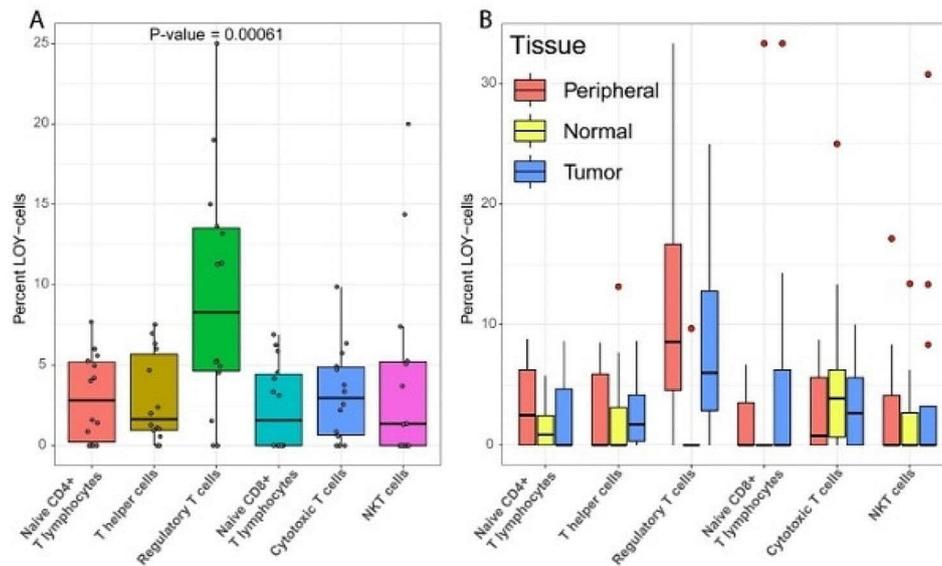
(scRNAseq) in multiple datasets, as well as T lymphocyte enrichment, to study how Y loss influences the developmental trajectory, distribution and gene expression of this rare leukocyte population.

## Results

### Distribution of LOY in T lymphocyte subsets

To investigate the distribution of LOY cells in CD4+ T lymphocytes we collected three public scRNAseq datasets containing fluorescence-activated cell sorted (FACS) T lymphocytes from cancer patients, with samples taken from tumour, healthy tissue and blood [22–24]. After standard pre-processing and clustering 21,709 cells remained, out of which 8,647 originated from male donors. Known marker genes were used to assign cell types (Supplementary Fig. 9A) to each cluster (Supplementary Fig. 9B) based on all 21,709 cells. The abundance of reads from the male specific region of the Y chromosome (MSY) was thereafter used to assign LOY status to each of the 8,647 cells from male donors. Due to the nature of the dataset, some cells from female donors also had MSY reads. Thus, the expression from MSY in female cells was considered as technical noise and used as a threshold when estimating the LOY status of single-cells from men. For each cell type the percentage of LOY-cells was calculated, showing LOY at marginal levels for the majority of the T lymphocytes. Strikingly, Tregs had a significant elevation of LOY with a close to three times higher occurrence of Y loss compared to other types of T lymphocytes (Wilcoxon signed-rank test:  $p=0.00061$ , Fig. 1A).

Given the observation that Tregs are the main carriers of LOY in the public datasets, we sought to investigate if Y loss might influence the distribution of T lymphocyte subtypes. First we used linear regression to test whether the fraction of Tregs (relative to the total number of CD4+ T lymphocytes) in each subject was associated with the percentage of LOY in Tregs. This analysis showed that subjects with higher levels of LOY in Tregs had a larger fraction of Tregs compared with other subjects (Coeff=2.622,  $p=0.033965$ ). Next, we evaluated this result in a model adjusting for relevant confounding factors such as dataset and individual. In support of the unadjusted model, a negative binomial model showed that the number of cells within each cell type was significantly associated with the level of LOY (GLM: LR=6.591, Df=1,  $p=0.01025$ ). Other significant confounders included cell type, as well as the interaction between Y loss and cell type (Supplementary Table 2). Finally, we also investigated how LOY varied between the sampled tissues (Fig. 1B). While LOY-cells were identified in all cell types and tissues, Tregs and CD8+ T lymphocytes with LOY were principally absent from normal tissue. To test the association between Y loss and sampled tissue,



**Fig. 1** (A) Boxplot showing the overall percentage of LOY in different T lymphocyte subsets, in each studied subject of the public datasets. Each dot corresponds to the LOY frequency of one patient, color-coded for each cell type. (B) Similar boxplot as in A, stratified by tissue of origin for T lymphocytes as indicated. Peripheral, Normal and Tumour indicates peripheral blood, normal tissue adjacent to tumour and tumour tissue, respectively. Note that the dots here marks outliers, and has therefore been coloured red to differentiate them from the dots in A

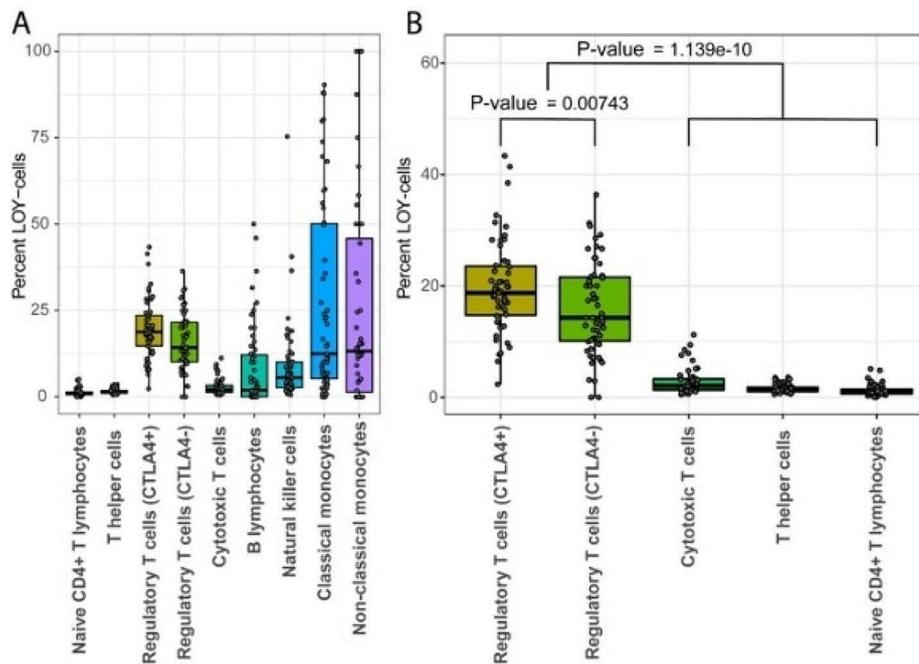
a quasibinomial model was used. This test confirmed that the level of LOY varied between tissues (GLM: LR=8.426, Df=2,  $p=0.014804$ ), as well as cell type, cancer type and the interactions between cell type, cancer type and tissue (Supplementary Table 3). Together, these results suggest that LOY in T lymphocytes might influence the distribution of T lymphocyte subsets. To further investigate a potential effect from Y loss on T lymphocyte development, we generated a larger scRNAseq dataset from an independent cohort as described below.

#### Validation of LOY distribution in an independent cohort

We performed scRNAseq on peripheral blood mononuclear cells (PBMCs), from freshly collected blood samples, from 55 males in the EpiHealth, UCAN and the UAD cohort (see methods). Enrichment for CD4+T lymphocytes was performed in 32 of these (from the EpiHealth and UCAN cohorts) prior to sequencing to achieve a larger fraction of Tregs in the studied cell population. After standard QC of the scRNAseq data, the final dataset consisted of all 55 samples and 213,619 single-cells, grouped into 27 clusters based on RNA expression profiles (Supplementary Fig. 10A). The major cell type in each cluster was identified combining two different machine learning approaches, CIPR and singleCellNet.

Both classification tools were trained on publicly available datasets (see methods) and used to predict the most likely cell type of each cluster. The final cell type classification was generated using the resulting predictions combined with manual curation based on the expression of known marker genes (Supplementary Fig. 10B). Next, LOY status was estimated for each cell separately, classifying cells without any MSY reads as LOY-cells. This method considers all reads within transcripts from MSY genes, unlike our previously published method, which only considered spliced RNA [12]. By applying a published benchmarking method for LOY scoring [25] that considers the overall MSY gene expression in each cluster, we found that our new method of LOY calling had considerably improved accuracy (Supplementary Figs. 1 & 2).

The frequency of LOY was thereafter characterized in the different types of studied leukocytes (Fig. 2A). In line with previous results, NK cells and monocytes exhibited the highest percentage of Y loss in individual samples [3, 12, 13]. However, Tregs had the highest median LOY value (17.91%) of all cell types in the validation dataset (Fig. 2A), replicating the observation from the public dataset (Fig. 1A). Furthermore, Tregs showed a significantly higher LOY fraction than other T cells (Wilcoxon



**Fig. 2** (A) Boxplot showing the percentage of LOY in the studied leukocytes of the validation dataset. Each dot represents the value of one sample. Median values are marked with black lines, where Tregs have the highest median of any studied cell type. (B) Subset of A, specifically the T lymphocytes

signed-rank test:  $p=1.139\text{e-}10$ , Fig. 2B) and was the only CD4+subset to harbour more than 5.12% LOY-cells, in any of the studied subject. Interestingly, the tools applied for cell clustering identified two separate Treg clusters (Supplementary Fig. 10). When further investigated, *CTLA4* expression was found as a major differentiating factor between the Treg clusters. The *CTLA4* gene is expressed on all mature Tregs and is a known regulator of Treg homeostasis [26, 27].

Overall, the 8,752 Tregs consisted of 6,975 CTLA4+ and 1,777 CTLA4- cells, and a significant difference in LOY level was found between the two Treg clusters (Wilcoxon signed-rank test:  $p=0.00743$ ). The higher frequency of LOY in CTLA4+ cells could suggest that Y loss is influencing the differentiation process of Tregs.

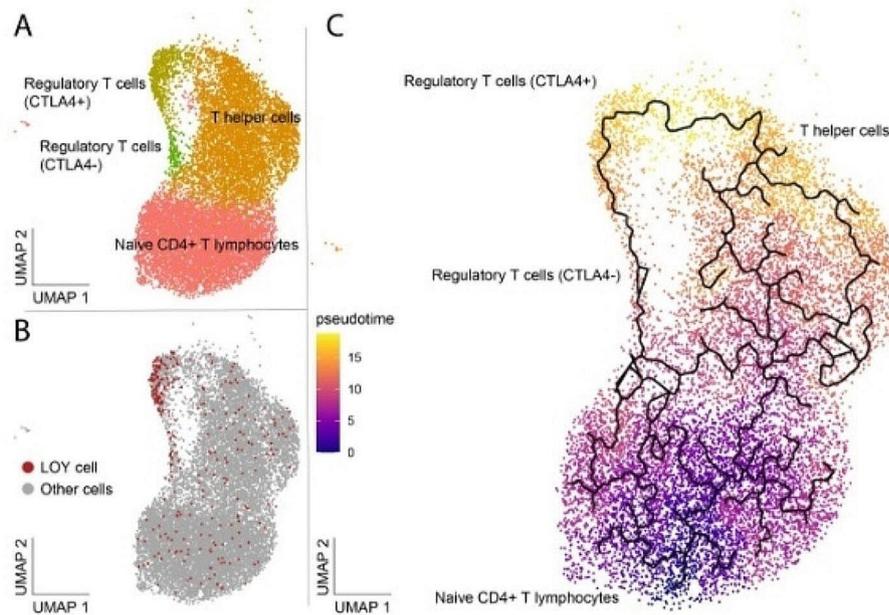
#### LOY as a determinant of CD4+ T lymphocyte cell fate

To replicate the finding from the public datasets that LOY could influence the differentiation of T lymphocytes towards a Treg phenotype, we used a similar approach in the validation dataset. First, linear regression was used to show that the level of LOY in Tregs was positively associated with the fraction of Tregs (Coeff=0.0394,  $p=0.00743$ ). Next, a quasibinomial model was used to establish this result while adjusting for confounders

(LR=5.63, Df=1  $p=0.0176167$ ). Significant confounders in the model was cell type, whether the sample was CD4+enriched and interactions between LOY and cell type, as well as cell type and CD4+enrichment (Supplementary Table 4). Overall, these results independently support the observation from the public datasets that LOY influence Treg abundance. Given this observation, we sought to investigate the hypothesis that Y loss could be affecting the development of CD4+ T lymphocytes by pushing them towards a Treg phenotype. Developmental trajectories for CD4+ T lymphocytes were estimated with pseudotime as a measurement for cell differentiation (Fig. 3). The most differentiated CD4+ cells comprised of CTLA4+ Tregs and adjacent T-helper cells, creating a trajectory suggesting a differentiation of naive T lymphocytes with Y loss into CTLA4+ Tregs via the CTLA4- subtype.

#### LOY associated transcriptional effects (LATE) in Tregs

Differential expression analysis was used to study changes in autosomal gene expression, referred to as LOY associated transcriptional effects (LATE), in Tregs with Y loss. For the CTLA4- subset of Tregs, only 23 LATE genes were identified (Supplementary Table 5). In contrast, analysis of the CTLA4+ subset of Tregs



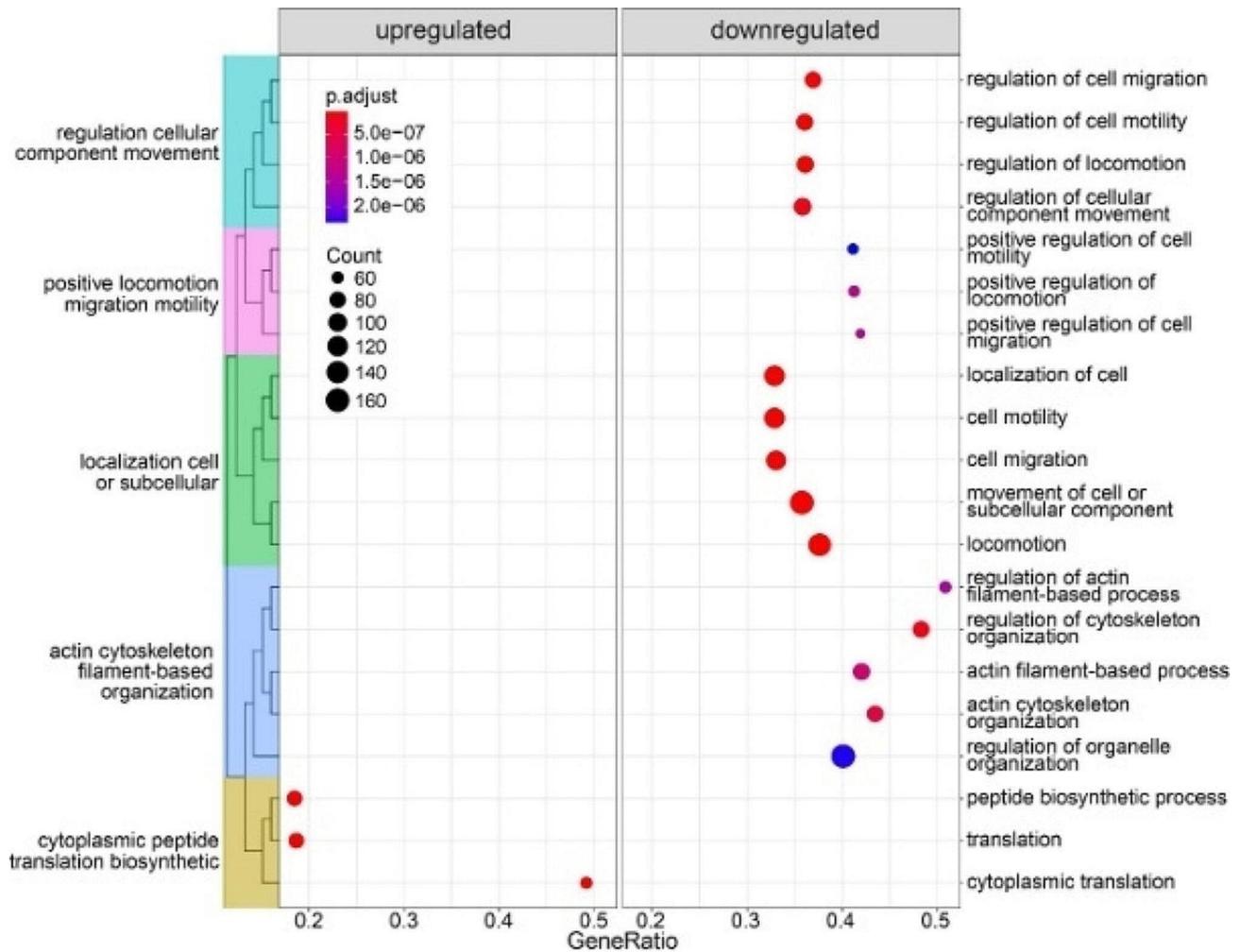
**Fig. 3** Illustration of the distribution of LOY in different types of CD4+T lymphocytes, as well as their developmental trajectories. The full UMAP can be seen in Supplementary Fig. 10B. **(A)** The distribution of the identified cell types. **(B)** The LOY status of the cells in **A**, with LOY-cells being marked red. **(C)** Developmental trajectory of CD4+T lymphocytes. Colour denotes pseudotime, with more developed cells being brighter. The lines indicate the suggested trajectories from naive to more differentiated CD4+ cells.

identified 465 LATE genes, most of which were autosomal genes (Supplementary Table 6). The list of differentially expressed genes contains many that are involved in the normal functions of immune cells, including *S100A11* (logFC = -0.25, adjusted p-value=8.9e-20), *ANXA1* (logFC = -0.18, adjusted p-value=4.6e-14), *TIGIT* (logFC=0.11, adjusted p-value=0.00011) and *FOXP3* (logFC=0.10, adjusted p-value=0.00015). Gene Set Enrichment Analysis (GSEA) found no gene sets for the CTLA4- Treg subset, while the CTLA4+subset yielded 41 significantly ( $p < 0.001$ , Supplementary Table 7) differentiated gene sets. The 20 most significant gene sets from the CTLA4+analysis were also grouped further into functional categories (Fig. 4, Supplementary Fig. 11). All significantly upregulated gene sets shared a core of upregulated ribosomal proteins (RPs). In contrast, the downregulated gene sets mainly consisted of genes involved in cell migration and locomotion.

## Discussion

Previous studies have established that LOY is associated with the distribution of different types of blood cells [10–12, 14]. Furthermore, it has been shown that the

type of leukocyte affected with LOY might be relevant for disease risks, with LOY in specifically CD4+T lymphocytes being associated with increased risk for prostate cancer [12]. However, LOY occurs at considerably lower frequency in CD4+T lymphocytes than other leukocytes [3, 12, 13]. Thus, we investigated here if Y loss disproportionately affects certain subsets of T lymphocytes. In both the public and validation datasets, we found that Tregs had significantly more LOY than any other studied CD4+T lymphocyte subset. Additionally, higher levels of LOY in Tregs was also positively associated with a higher frequency of Tregs compared with other CD4+cells. This suggests that Y loss might impact the distribution of T lymphocytes, by pushing naive T lymphocytes towards a Treg phenotype. Alternatively, CD4+T lymphocytes affected by LOY could undergo apoptosis, with the enrichment for LOY Tregs being due to a lower susceptibility for this apoptosis. However, analysis of developmental trajectories in the validation dataset support the first hypothesis; since it predicts trajectories culminating in the Treg cluster, which has the highest frequency of LOY-cells. Overall, our data suggest that Y loss influence



**Fig. 4** The top 20 categories from the gene set enrichment analysis. Categories within the left and right frames were upregulated and downregulated, respectively. The coloured clusters to the left indicate the same branch of the gene ontology (GO) tree. Dot size and colour indicates gene count and P-value, respectively

the differentiation of CD4<sup>+</sup>T lymphocytes, resulting in an enrichment of Tregs with LOY.

LATE analysis in CTLA4<sup>+</sup>Tregs identified 465 dysregulated genes, and the GSEA identified 41 gene sets describing two major effects associated with Y loss. First, the increased expression of RPs identified in LOY Tregs indicates that these cells are upregulating production of ribosomes. Interestingly, losing the Y chromosome involves losing *RPS4Y1*, coding for a ribosomal subunit, and located in the MSY. Its X chromosome homolog, *RPS4X*, escapes X-inactivation, suggesting that the expression of two copies is necessary to maintain dosage [28]. Deletion of ribosomal proteins can activate the

mTOR pathway and disrupts ribosomal assembly, resulting in RP upregulation [29] and transcriptional dysregulation [30]. Thus, the loss of *RPS4Y1* could explain the observed upregulation of RPs, and might contribute to 465 LATE genes found in LOY Tregs. The second major effect identified by the GSEA was several gene sets linked with cell motility. This could be attributed to *CD99*, a gene located in the pseudoautosomal region that will have one copy lost as an effect of Y loss [31]. Studies have previously reported downregulation of *CD99* in LOY-cells, as well as a decreased cell surface abundance of CD99, as an effect of Y loss [12, 13]. While a decreased motility of leukocytes with LOY could impair their normal immune

functions, it might also influence the varying distribution of LOY-cells in different tissues observed in the public dataset. Here, normal tissue neighbouring tumours was depleted of Tregs and CD8+T lymphocytes with Y loss, compared with high levels of LOY found in blood and tumour tissue. Tumours are known to recruit suppressive immune cells [19, 20], and it is therefore possible that Tregs with Y loss might be concentrated in the tumour microenvironment due to decreased mobility.

In addition to major effects identified by GSEA, certain genes dysregulated in Tregs with LOY should be highlighted. First, both *S100A11* and *ANXA1* genes are downregulated as an effect of Y loss. They encode proteins that forms a complex capable of regulating the EGFR pathway [32–34], with *S100A11* also being a part of the TGF-beta signalling pathway [35]. In specifically T lymphocytes, *ANXA1* is conversely an important modulator of proinflammatory functions [36], with evidence of *ANXA1* decreasing the risk of atherosclerosis in humans [37]. *ANXA1* knockout in mouse models lead to chronic inflammation, including lung fibrosis, sepsis, rheumatoid arthritis and atherosclerotic lesion formation [38]. Thus, downregulation of *ANXA1* in Tregs with Y loss could severely limit their normal functions, and in extension inhibit an inflammatory response.

Another interesting gene that is upregulated in Tregs with LOY is *TIGIT*, an immunosuppressive receptor found on tumour-infiltrating NK cells, CD8+, CD4+ and regulatory T cells, with highest abundance in the latter [39]. The *TIGIT* receptor inhibits immune function by binding with higher affinity to CD155 and CD112 than their usual receptor CD226 [39, 40]. While the CD226 binding would enhance T lymphocyte and NK cell activation, *TIGIT* binding instead actively suppress immune functions of these cells [41–43]. However, in Tregs, *TIGIT* is a marker for stability, promoting their immunosuppressive functions further [44]. Additionally, an increased *TIGIT* to CD226 ratio in the tumour microenvironment has been associated with a higher frequency of activated Tregs, as well as an unfavourable prognosis [45]. Since differential expression was analysed in the PBMC based validation dataset, it does not present tumour microenvironment and cancer type specific effects, which could be interesting aspects in future studies.

Finally, *FOXP3* was upregulated in LOY Tregs, a major determinant for Treg development and their immunosuppressive activity [46]. Loss of function mutations in *FOXP3* has previously been associated with hyperactive T lymphocytes, as well as fatal immunodysregulation [47]. *FOXP3* exerts genome wide regulation of gene expression [48], including promotion of *CTLA4* [49], which might be related to the increased level of LOY in the *CTLA4*+Treg subset observed here. Since *FOXP3*

drives Treg development [46], its upregulation in LOY-cells provides a possible mechanism by which Y loss could influence CD4+T lymphocyte development.

## Conclusions

Here we suggest a possible mechanism to help explain why men with hematopoietic Y loss may have an increased risk of tumour development in other organs. Taken together, our data indicate that LOY could drive the development of CD4+T lymphocytes towards a regulatory phenotype, leading to enrichment of Tregs with Y loss. Differential expression analysis further highlight genes involved in the immunosuppressive functions of these regulatory cells, potentially linked with the increased vulnerability for cancer previously observed in men affected with LOY.

Methods.

## Collection and sequencing

The patient cohort selected for enrichment of CD4+leukocytes included 30 male participants from the Epidemiology for health study (EpiHealth) and 2 males from the Uppsala-Umeå Comprehensive Cancer Consortium (UCAN). 32 ml of blood was collected into four BD Vacutainer® CPT™ Mononuclear Cell Preparation Tubes (BD), and PBMCs were isolated following the manufacturer's instructions. The PBMCs were then washed with PBS and cell number and viability were estimated with *EVE*™ Automated Cell Counter (NanoEnTek) using trypan blue. CD4+T lymphocytes were enriched from the PBMCs using CD4+T Cell Isolation Kit human (Miltenyi Biotec) according to the manufacturer's instructions. Enriched CD4+T-cells were then diluted to a concentration of  $10^6$  cells/ml in PBS with 0.04% BSA with a cell viability of >90%. scRNAseq libraries of CD4+T-cells were generated using Chromium Next GEM Single Cell 3' Reagent kit v3.1 (10x Genomics) according to the manufacturer's instructions. The single-cell libraries were then sequenced using the NovaSeq 6000 and v1.5 sequencing chemistry (Illumina Inc.). The single-cell library preparation and sequencing were performed at the Science for Life technology platform SNP&SEQ, Uppsala University, Sweden. Additionally, 23 scRNAseq datasets derived from PBMCs from the Uppsala Alzheimer's Disease cohort (UAD) were used as validation. Sample preparation was the same as above except for the CD4+T Cell enrichment. Overview of participants and their clinical characteristics can be seen in Supplementary Table 1.

## Pre-processing, mapping and LOY cell identification

Each sequenced sample was mapped using the Cell Ranger pipeline (v. 6.0 10X Genomics) and standard settings. Following this, the *velocyto* software was used, counting reads from expressed transcripts and also

counting intronic reads as well as reads in the untranslated regions. Generated count matrices from both software were read into an R environment for further study. To identify cells with LOY, all generated count matrices were used and cells showing no reads mapping to the male specific Y were considered to be LOY cells. In contrast to previous methods, this included any MSY reads identified by velocity as well as cell ranger. This was benchmarked using an established method recommending that each cluster has a score of at least 250 expression from MSY to call LOY-status [25]. The new approach scored higher than 250 in all clusters (Supplementary Fig. 1'), which would not be the case with previous methods (Supplementary Fig. 2).

#### Data harmonization and quality control

Unless specified otherwise, the following analysis steps were performed in R version 4.0.4, using Seurat version 4.0.1, on the UPPMAX Bianca computational cluster. Two criteria were used to exclude low quality cells, the number of expressed features and the percent mitochondrial reads. Cells were required to have more than 500, but less than 3000 expressed features, as well as less than 13% of all reads being mitochondrial. These thresholds were chosen based on the variable distribution (Supplementary Fig. 3). The count matrix was thereafter normalized using SCTransform (version 0.3.2), regressing out effects based on percent mitochondrial reads. The samples were thereafter integrated, based on 3000 features and one sample (SF-2212-EPH001) as reference. The integration functions were chosen to account for the SCTransform. In addition to SCTransform, log-based normalization was also done using Seurats NormalizeData function with default settings. The log-based normalization is preferred when investigating differentially expressed genes, as advised by the team behind Seurat. After calculating principal components, batch effects introduced by sampling and sorting were removed using Harmony (version 1.0). The number of harmonized principal components to use for clustering was thereafter chosen based on an Elbowplot (Supplementary Fig. 4). Clustering was performed on the 22 first principle components with 0.9 resolution.

#### Cell type identification

The classification of cell type identity was guided by two tools, CIPR (version 0.1.0) and singleCellNet (version 0.1.0). CIPR was run on marker genes, calculated using the FindAllMarkers function on log-normalized data. The Presorted PBMC single-cell RNAseq dataset, hsrnaseq, was used as reference. The resulting CIPR classification can be seen in Supplementary Fig. 5A. SingleCellNet was trained using data from Zheng et al. [50], available through the 10X Genomics Datasets

database. The training dataset was filtered ( $200 < \text{number of features} < 1500$  &  $\text{percent mitochondrial} < 5\%$ ) and normalized according to the main dataset. It was then trained ( $n\text{TopGenes}=10$ ,  $n\text{Rand}=70$ ,  $n\text{Trees}=1000$ ,  $n\text{TopGenePairs}=25$ ) and prediction scores tested ( $n\text{rand}=50$ ). Thereafter, singleCellNet classification was run ( $n\text{qRand}=50$ ) on the log-normalized main data. The singleCellNet classification can be seen in Supplementary Fig. 5B, where the most commonly predicted cell type per cluster was used as identity for the entire cluster. The cell identities suggested by each tool were used to guide cell type classification, which ultimately was decided based on the expression of known marker genes in each cluster. These known marker genes included *CD4*, *FHIT* and *CCR7* for Naïve CD4+T lymphocytes, with *FHIT* and *CCR7* negative as Helper T cells. *CD4*, *FOXP3*, *IL2RA* and *TIGIT* for Tregs, which were further separated by *CTLA4*. *CD8* for Cytotoxic T cells, while *NKG7* and *GZMB* indicated NK cells. Classical monocytes were identified by *FCN1* and *CD14*, with non-classical monocytes defined by the addition of *FCGR3A*. B-lymphocytes by *CD19* and *VPREB3*. Additional clusters, exhibiting gene expression profiles not indicating any of the above marker genes, were classified as unidentified.

#### Differential expression analysis

Genes differentially expressed due to LOY were found with the Limma-trend algorithm. This was done per cell type, and included genes expressed in at least 10% of cells in the studied cell type. Using the Limma R package (version 3.46.0), the model matrix was defined with LOY status and sample origin specified as covariates. The model was thereafter fitted on the Log-normalized expression data with default settings, as well as LOY status set as the coefficient of interest and Benjamini-Hochberg as p-value adjustment method.

#### Gene set enrichment analysis

Fold changes for all genes, calculated using the limma package (see differential expression analysis), was collected. Genes on the male specific region of the Y chromosome were removed. After this, clusterProfiler (version 4.2.2) was used with fold changes for each gene, tested by the limma package, as a metric to calculate the enrichment of all gene sets present in the "Biological process" category of the gene ontology resource (<http://geneontology.org/>). Clustering for categories was performed using the pairwise\_termsim function from the enrichplot R package (version 1.13.1.992).

#### Developmental trajectories

The cells defined as Naïve CD4+T lymphocytes, Helper T cells and Tregs were selected to estimate developmental trajectories. To avoid an issue with overfitting,

the number of cells was randomly decreased by a factor of 10 using the sample function in R. The seed was set to 14 for this step. The SeuratWrapper package (version 0.3.0) was thereafter used in R (version 4.1.3) to transform the sampled SeuratObject into a CellDataSet object used by Monocle3 (version 1.0.0), which was used to predict developmental trajectories. Monocle3 was run on the UMAP previously constructed, using the Louvain clustering method. The cluster designated as Y\_73 was thereafter used as the root. Prior to running the developmental trajectory analysis, the random generator seed was set to 1477.

### Public dataset

The three public datasets were collected via their corresponding GEO catalogues, GSE98638, GSE99254 and GSE108989. They were processed similarly to as described above, with the exception of the steps described here. The analysis was performed in R version 4.1.3 with Seurat version 4.1.0. Any step prior to integration was performed independently for each dataset. Firstly, low quality cells were filtered based on the number of expressed features (nFeats) and UMI count (nUMI). nFeats was used instead of percent mitochondrial read as this was not available for the public datasets. The thresholds were chosen based on the corresponding distribution (Supplementary Fig. 6) for liver cancer ( $2300 < nFeats < 4400$ ;  $3e5 < nUMI < 1.2e6$ ), lung cancer ( $1800 < nFeats < 5200$ ;  $2e5 < nUMI < 1e6$ ) and colorectal cancer ( $1800 < nFeats < 5200$ ;  $2e5 < nUMI < 1e6$ ). Metadata was thereafter collected for the cells that passed filtering via the identifier assigned to each cell by the original authors. Except for the numbers identifying the patient of origin for each cell, combinations of letters could be used to discern the sampled tissue and sorting. In addition to P, T and N denoting peripheral blood, tumor tissue and normal adjacent tissue, TR, TC and TH identified Tregs, cytotoxic T cells and T helper cells, respectively. Other identifiers were not present in all three datasets; these cells were therefore excluded. To classify LOY-status, the sex of each original sample was identified by comparing the expression of MSY genes. The list of Y located genes was collected using the BioMart package (version 2.40.0) on Ensembl (version 99). While the sex of each patient was clear, some female cells still expressed MSY genes. Considering the female MSY expression background noise, a threshold was created at the 95th quantile of total MSY expression in specifically female cells with any MSY expression. Additionally, most female MSY expression was from a single MSY gene. Thus, any male cells with a total MSY expression less than this threshold, as well as expression from only one or less MSY gene, were classified as LOY (Supplementary Fig. 7). After normalization using SCTransform (version 0.3.3). The three datasets

were integrated, followed by the calculation of principal components. Harmony was used to remove batch effects attributed to patients and tissue of origin.

The first 15 dimensions from Harmony, chosen based on elbow plot (Supplementary Fig. 8), were finally used to cluster the cells to a resolution of 0.6. Due to the poor performance on this dataset by both cell type classification algorithms, the cell type of each cluster was manually designated based on only the expression of known marker genes.

### Statistics

To test if LOY was more common in Tregs than other T lymphocytes, Wilcoxon signed-rank test was used to compare the difference between LOY percentage values from the sample. For the validation dataset this was first done between Tregs and other T lymphocytes to reduce the number of tests performed, thereafter testing the difference between the Treg subsets. When comparing the Treg versus other CD4+ T lymphocytes abundance, non-parametric linear regression was used via the mblm package (version 0.12.1). A non-parametric model was necessary as the variables were not normally distributed. Further models to test the association between LOY-level and other factors such as cell type and tissue was run either as an quasibinomial model with R's glm function or a negative binomial model run with the MASS package (version 7.3–55). The negative binomial model was used for the LOY to cell type test in the public datasets to account for additional data structures and sorting. Quasi models was also necessary to handle high levels of residual deviance. The produced models were run as a type 3 ANOVA using the car package (version 3.0–10 and version 3.0–12 for the validation and public datasets, respectively), with contrasts defined as options(contrasts=c("contr.sum", "contr.poly")). See Supplementary Tables 2–4 for full models.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10168-7>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

### Acknowledgements

We would like to thank the study nurse Malin Edén at the memory and geriatric clinic, academic hospital for help with the sample collection, as well as the authors of the publically available datasets.

### Author contributions

J.M., J.H., H.D., P.O., M.D., E.R-B., J.P.D. and L.A.F designed the study. J.H., J.P.D. and L.A.F obtained the funding. H.D., B.B-O., P.O., M.D. and E.R-B. performed the experiments. J.M., J.H., J.B., A.L., A.Z. and L.A.F analysed the data and interpreted the results. H.D., B.B-O. and J.P.D. contributed to sample collection.

J.M., J.H. H.D., J.B. and L.A.F. wrote the first version of the paper. All authors contributed to the editing of, and approved, the final manuscript.

### Funding

This result is part of a project L.A.F. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreements No. 679744 and 101001789), Additional funding supporting the research to L.A.F.: Swedish Research Council (2017–03762 and 2022–03452), The Swedish Cancer Society (20-1004), Kjell och Märta Beijers Stiftelse, and Konung Gustav V och Drottning Victorias Stiftelse. This study was supported by grants from Swedish Cancer Society, Swedish Research Council (grant number 2020–02010), Swedish Heart-Lung Foundation (grant number 20210051), Hjärnfonden and the Foundation for Polish Science under the International Research Agendas Programme (grant number MAB/2018/6; co-financed by the European Union under the European Regional Development Fund) to J.P.D. J.H. received funding from Marcus Borgströms stiftelse. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The computations and data handling was enabled by resources in project sens2017-134 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022–06725.

Open access funding provided by Uppsala University.

### Data availability

Public datasets are available through GEO catalogues GSE98638, GSE99254 and GSE108989. The dataset generated for the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Informed consent was obtained from all the participants, or from next of kin. The research was approved by the local research ethics committee in Uppsala, Sweden Dnr. 2015/458 (EpiHealth cohort), amendment Dnr. 2015/458/2 (UCAN cohort), Dnr. 2015/092 (UAD cohort).

#### Consent for publication

Not applicable.

#### Competing interests

J.P.D. and L.A.F. are cofounders and shareholders in Cray Innovation AB. The remaining authors declare no competing interest.

Received: 27 July 2023 / Accepted: 28 February 2024

Published online: 05 March 2024

### References

- Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease — clones picking up speed. *Nat Rev Genet*. 2017;18:128–42. <https://doi.org/10.1038/nrg.2016.145>.
- Zhou W, et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat Genet*. 2016;48:563–8. <https://doi.org/10.1038/ng.3545>.
- Thompson DJ, et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature*. 2019;575:652–7. <https://doi.org/10.1038/s41586-019-1765-3>.
- Forsberg LA, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014;46:624–8. <https://doi.org/10.1038/ng.2966>.
- Lofffield E, et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci Rep*. 2018;8:12316. <https://doi.org/10.1038/s41598-018-30759-1>.
- Haitjema S, et al. Loss of Y chromosome in blood is Associated with Major Cardiovascular events during Follow-Up in men after Carotid Endarterectomy. *Circ Cardiovasc Genet*. 2017;10:e001544. <https://doi.org/10.1161/CIRCGENETICS.116.001544>.
- Sano S, et al. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science*. 2022;377:292–7. <https://doi.org/10.1126/science.abn3100>.
- Ganster C, et al. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. *Genes Chromosomes Cancer*. 2015;54:717–24. <https://doi.org/10.1002/gcc.22282>.
- Dumanski JP, et al. Mosaic loss of chromosome Y in blood is Associated with Alzheimer Disease. *Am J Hum Genet*. 2016;98:1208–19. <https://doi.org/10.1016/j.ajhg.2016.05.014>.
- Terao C, et al. GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nat Commun*. 2019;10:4719. <https://doi.org/10.1038/s41467-019-12705-5>.
- Lin SH, et al. Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK Biobank men. *Sci Rep*. 2020;10:3655. <https://doi.org/10.1038/s41598-020-59963-8>.
- Dumanski JP, et al. Immune cells lacking Y chromosome show dysregulation of autosomal gene expression. *Cell Mol Life Sci*. 2021;78:4019–33. <https://doi.org/10.1007/s00018-021-03822-w>.
- Mattisson J, et al. Leukocytes with chromosome Y loss have reduced abundance of the cell surface immunoprotein CD99. *Sci Rep*. 2021;11:15160. <https://doi.org/10.1038/s41598-021-94588-5>.
- Bruhn-Olszewska B, et al. Loss of Y in leukocytes as a risk factor for critical COVID-19 in men. *Genome Med*. 2022;14:139. <https://doi.org/10.1186/s13073-022-01144-5>.
- Baron U, et al. DNA demethylation in the human FOXP3 locus discriminates regulatory T cells from activated FOXP3+ conventional T cells. *Eur J Immunol*. 2007;37:2378–89. <https://doi.org/10.1002/eji.200737594>.
- Dieckmann D, Plottner H, Berchtold S, Berger T, Schuler G. Ex vivo isolation and characterization of Cd4+ Cd25+ T cells with Regulatory properties from Human Blood. *J Exp Med*. 2001;193:1303–10. <https://doi.org/10.1084/jem.193.11.1303>.
- Sakaguchi S, et al. Regulatory T cells and human disease. *Annu Rev Immunol*. 2020;38:541–66. <https://doi.org/10.1146/annurev-immunol-042718-041717>.
- Rocamora-Reverte L, Melzer FL, Würzner R, Weinberger B. The Complex Role of Regulatory T Cells in immunity and aging. *Front Immunol*. 2020;11:616949. <https://doi.org/10.3389/fimmu.2020.616949>.
- Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. *Cell Res*. 2017;27:109–18. <https://doi.org/10.1038/cr.2016.151>.
- Ménétrier-Caux C, et al. Targeting regulatory T cells. *Target Oncol*. 2012;7:15–28. <https://doi.org/10.1007/s11523-012-0208-y>.
- Abdel-Hafiz HA, et al. Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature*. 2023;619:624–31. <https://doi.org/10.1038/s41586-023-06234-x>.
- Zheng C, et al. Landscape of infiltrating T cells in Liver Cancer revealed by single-cell sequencing. *Cell*. 2017;169:1342–1356e1316. <https://doi.org/10.1016/j.cell.2017.05.035>.
- Guo X, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med*. 2018;24:978–85. <https://doi.org/10.1038/s41591-018-0045-3>.
- Zhang L, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. 2018;564:268–72. <https://doi.org/10.1038/s41586-018-0694-x>.
- Vermeulen MC, Pearce R, Young-Pearse T, Mostafavi S. Mosaic loss of chromosome Y in aged human microglia. *Genome Res*. 2022;32:1795–807. <https://doi.org/10.1101/gr.276409.121>.
- Barnes MJ, et al. CTLA-4 promotes Foxp3 induction and regulatory T cell accumulation in the intestinal lamina propria. *Mucosal Immunol*. 2013;6:324–34. <https://doi.org/10.1038/mi.2012.75>.
- Zhao H, Liao X, Kang Y, Tregs. Where we are and what comes next? *Front Immunol*. 2017;8:1578. <https://doi.org/10.3389/fimmu.2017.01578>.
- Peeters SB, et al. How do genes that escape from X-chromosome inactivation contribute to Turner syndrome? *Am J Med Genet C Semin Med Genet*. 2019;181:28–35. <https://doi.org/10.1002/ajmg.c.31672>.
- Goudarzi KM, Lindström MS. Role of ribosomal protein mutations in tumor development (review). *Int J Oncol*. 2016;48:1313–24. <https://doi.org/10.3892/ijo.2016.3387>.
- Genuth NR, Barna M. The Discovery of Ribosome Heterogeneity and its implications for Gene Regulation and Organismal Life. *Mol Cell*. 2018;71:364–74. <https://doi.org/10.1016/j.molcel.2018.07.018>.
- Goodfellow P, Pym B, Mohandas T, Shapiro LJ. The cell surface antigen locus, MIC2X, escapes X-inactivation. *Am J Hum Genet*. 1984;36:777–82.

32. Zhang L, Zhu T, Miao H, Liang B. The calcium binding protein S100A11 and its roles in diseases. *Front Cell Dev Biol.* 2021;9:693262. <https://doi.org/10.3389/fcell.2021.693262>.
33. Poeter M, et al. Disruption of the annexin A1/S100A11 complex increases the migration and clonogenic growth by dysregulating epithelial growth factor (EGF) signaling. *Biochim Biophys Acta.* 2013;1833:1700–11. <https://doi.org/10.1016/j.bbamcr.2012.12.006>.
34. de Graauw M, et al. Annexin A1 regulates TGF-beta signaling and promotes metastasis formation of basal-like breast cancer cells. *Proc Natl Acad Sci U S A.* 2010;107:6340–5. <https://doi.org/10.1073/pnas.0913360107>.
35. He H, Li J, Weng S, Li M, Yu Y. S100A11: diverse function and Pathology corresponding to different target proteins. *Cell Biochem Biophys.* 2009;55:117. <https://doi.org/10.1007/s12013-009-9061-8>.
36. Yang YH, et al. Deficiency of annexin A1 in CD4+T cells exacerbates T cell-dependent inflammation. *J Immunol.* 2013;190:997–1007. <https://doi.org/10.4049/jimmunol.1202236>.
37. de Jong R, Leoni G, Drechsler M, Soehnlein O. The advantageous role of annexin A1 in cardiovascular disease. *Cell Adh Migr.* 2017;11:261–74. <https://doi.org/10.1080/19336918.2016.1259059>.
38. Grewal T, Wason SJ, Enrich C, Rentero C. Annexins— insights from knockout mice. *Biol Chem.* 2016;397:1031–53. <https://doi.org/10.1515/hsz-2016-0168>.
39. Ge Z, Peppelenbosch MP, Sprengers D, Kwekkeboom J. TIGIT, the next step towards successful combination Immune Checkpoint Therapy in Cancer. *Front Immunol.* 2021;12:699895. <https://doi.org/10.3389/fimmu.2021.699895>.
40. Huang Z, Qi G, Miller JS, Zheng SG. CD226: an emerging role in Immunologic diseases. *Front Cell Dev Biol.* 2020;8:564. <https://doi.org/10.3389/fcell.2020.00564>.
41. Yu X, et al. The surface protein TIGIT suppresses T cell activation by promoting the generation of mature immunoregulatory dendritic cells. *Nat Immunol.* 2009;10:48–57. <https://doi.org/10.1038/ni.1674>.
42. Joller N, et al. Cutting edge: TIGIT has T cell-intrinsic inhibitory functions. *J Immunol.* 2011;186:1338–42. <https://doi.org/10.4049/jimmunol.1003081>.
43. Lozano E, Dominguez-Villar M, Kuchroo V, Hafler DA. The TIGIT/CD226 axis regulates human T cell function. *J Immunol.* 2012;188:3869–75. <https://doi.org/10.4049/jimmunol.1103627>.
44. Kurtulus S, et al. TIGIT predominantly regulates the immune response via regulatory T cells. *J Clin Invest.* 2015;125:4053–62. <https://doi.org/10.1172/jci81187>.
45. Fourcade J, et al. CD226 opposes TIGIT to disrupt Tregs in melanoma. *JCI Insight.* 2018;3. <https://doi.org/10.1172/jci.insight.121157>.
46. Lu L, Barbi J, Pan F. The regulation of immune tolerance by FOXP3. *Nat Rev Immunol.* 2017;17:703–17. <https://doi.org/10.1038/nri.2017.75>.
47. Bennett CL, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet.* 2001;27:20–1. <https://doi.org/10.1038/83713>.
48. Zheng Y, et al. Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature.* 2007;445:936–40. <https://doi.org/10.1038/nature05563>.
49. Chen C, Rowell EA, Thomas RM, Hancock WW, Wells AD. Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *J Biol Chem.* 2006;281:36828–34. <https://doi.org/10.1074/jbc.M608848200>.
50. Zheng GXY, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049. <https://doi.org/10.1038/ncomms14049>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.