Expression of SARS-CoV-2 entry factors in lung epithelial stem cells and its potential implications for COVID-19

Anna A. Valyaeva1, Anastasia A. Zharikova1,2, Artem S. Kasianov2, Yegor S. Vassetzky3,4,5,6 & Eugene V. Sheval1,5,6,

SARS-CoV-2 can infiltrate the lower respiratory tract, resulting in severe respiratory failure and a high death rate. Normally, the airway and alveolar epithelium can be rapidly reconstituted by multipotent stem cells after episodes of infection. Here, we analyzed published RNA-seq datasets and demonstrated that cells of four different lung epithelial stem cell types express SARS-CoV-2 entry factors, including Ace2. Thus, stem cells can be potentially infected by SARS-CoV-2, which may lead to defects in regeneration capacity partially accounting for the severity of SARS-CoV-2 infection and its consequences.

The emergence of the novel SARS coronavirus 2 (SARS-CoV-2), which causes the respiratory disease COVID-19, poses an important threat to global public health. In contrast to endemic human coronaviruses, SARS-CoV-2 can infiltrate the lower respiratory tract, resulting in severe respiratory failure and a high death rate1,2. Infection of epithelial cells of the two major compartments in the lungs, the conducting airways and the gas-exchanging alveoli, seems to be a main cause of COVID-19 development3. Both the airway and alveolar epithelia are characterized by relatively slow renewal but can be rapidly reconstituted by multipotent stem cells after episodes of infection, inflammation or injury, which are commonly observed in respiratory diseases4–6.

To date, several types of epithelial stem cells in the distal airways and pulmonary alveoli have been identified in mice. Basal cells in the proximal airways are the major stem cell population that self-renew and, when necessary, give rise to multiple cell types, such as secretory club cells and ciliated cells7–10. The pulmonary alveolar epithelium is composed of two types of differentiated epithelial cells: alveolar type 1 (AT1) cells, which mediate gas exchange, and alveolar type 2 (AT2) cells, which secrete surfactant. A subpopulation of AT2 cells serves as alveolar stem cells that can differentiate into both AT1 and AT2 cells11,12. In addition, lung injury activates specialized epithelial stem cells residing in the distal airways to reestablish alveolar barrier function. A number of distinct, small populations of stem cells have been reported to contribute to regeneration after lung injury: variant club/secretory cells (v-club cells), bronchioalveolar stem cells (BASCs), H2-K1high club cell-like stem cells and p63+ basal cells6.

Infection of epithelial stem cells can potentially lead to defects in lung regeneration capacity. Analysis of the expression of viral entry factors helps to identify human cells that can be infected by SARS-CoV-2. Cellular entry of coronaviruses depends on binding of the viral spike (S) proteins to cellular receptors and on S protein priming by host cell proteases. It was demonstrated that angiotensin-converting enzyme 2 (ACE2) is the entry receptor for SARS-CoV-21,13,14, and TMPRSS2 and FURIN are major proteases for S protein priming13,15. Thus, (co)expression of ACE2, TMPRSS2 and FURIN is a convenient marker of cells that can be potentially infected by SARS-CoV-216.

Additional proteases potentially involved in SARS-CoV-2 priming include ANPEP, used by HCoV-229E17, and DPP4, used by MERS-CoV18. However, no data have been published on this subject for SARS-CoV-2.
Expression of SARS-CoV-2 entry factors can be analyzed using publicly available RNA-seq datasets. These factors, including ACE2, are highly expressed in nasal epithelial cells, but ACE2 expression on the cells of conducting airways and lung parenchyma is substantially lower\(^{23,20}\). SARS-CoV-2 entry factors are expressed in secretory and ciliated cells of the conducting airways\(^{23,19}\). In the alveolar epithelium, ACE2 expression is found only in a small subset (1–7%) of AT2 cells\(^{19,21,22}\), although the severity of the disease suggests a more widespread distribution.

Infection of epithelial stem cells can lead to defects in lung regeneration, but the published data on the expression of SARS-CoV-2 entry factors in stem cells are restricted to basal cells of the alveolar epithelium. Here, to determine whether lung stem cells can be infected by SARS-CoV-2, we analyzed the expression of SARS-CoV-2 entry factors in different epithelial stem cells using publicly available RNA-seq data. Because of the limited data on lung stem cells in humans we conducted this study on data obtained from mice.

**Results**

Airway epithelial cell types include basal cells, secretory club cells and ciliated cells, as well as several rare cell types—neuroendocrine, goblet and tuft cells and ionocytes. Basal cells are a heterogeneous population of stem cells of the conducting airways that can self-renew and differentiate into both secretory and ciliated epithelial lineages.

Recent studies in which publicly available scRNA-seq data were reanalyzed to detect ACE2-expressing cells demonstrated that only a small subpopulation of basal cells express ACE2 and other SARS-CoV-2 entry factors; ACE2 expression increased upon differentiation to secretory club cells\(^9\). We proposed that the expression of ACE2 and other SARS-CoV-2 entry factors might be underestimated due to the presence of dropout events, i.e., a gene is detected in one cell but is not detected in another cell, usually due to extremely low mRNA input and/or the stochastic nature of gene expression\(^26–29\). We therefore compared datasets obtained by Montoro and coauthors using two different methods, 3′ single-cell RNA-seq (scRNA-seq) and full-length scRNA-seq\(^1\). Full-length scRNA-seq dataset has a higher number of reads per cell for each gene than 3′ scRNA-seq dataset (396,000 reads/cell and 23,000 reads/cell, respectively\(^9\)), which can lead to a much more precise estimation of the number of cells expressing SARS-CoV-2 entry factors.

We analyzed the expression of SARS-CoV-2 entry factors only in basal, ciliated and club cells because the number of other cells in the full-length scRNA-seq dataset was extremely low. The proportion of cells expressing Ace2 and the priming proteases was substantially higher in the full-length scRNA-seq data (Fig. 1A). For example, the proportions of Ace2+ basal cells were 0.60% and 9.38% in the 3′ scRNA-seq and full-length scRNA-seq datasets, respectively. This indicates that the population of epithelial cells potentially sensitive to SARS-CoV-2 infection is underestimated by 3′ scRNA-seq. The expression patterns of different genes by 3′ scRNA-seq and full-length scRNA-seq were similar for highly expressed genes (e.g., Tmprss2), although the detected expression levels of genes with relatively low mean expression levels were substantially higher for 3′ scRNA-seq (Fig. 1B). The proportions of cells with greater-than-zero expression of different genes were extremely variable between the two datasets, particularly for genes with low and moderate expression levels; thus the dropout effect was higher for these genes (Fig. 1C). Thus, the number of cells expressing SARS-CoV-2 entry factors obtained using 3′ scRNA-seq and similar methods used in cell atlas projects could be underestimated and should be carefully interpreted.

Only one type of lung stem cell—the basal cells of the conducting airways—has been extensively studied in both mice and humans. For the other types of stem cells, quality RNA-seq data can only be found for mice. However, the patterns of SARS-CoV-2 entry factor expression on different epithelial cells of mouse and human airways\(^9\) are highly similar (Supplementary Information Fig. 1), indicating that datasets from mice can be used for analysis of SARS-CoV-2 entry factors.

AT2 cells serve as alveolar stem cells and can differentiate into AT1 cells during alveolar homeostasis and postinjury repair\(^21,22\), but a small subpopulation of AT1 cells retains cellular plasticity and can transdifferentiate into AT2 cells, maintaining tissue integrity during alveolar regeneration\(^23,34\). These cells are characterized by the Hopx+/Lgfbp2+ phenotype, but they do not form separated clusters in t-distributed stochastic neighbor embedding (t-SNE) plots. We analyzed the expression of SARS-CoV-2 entry factors in Lgfbp2 and Lgfbp2+ AT1 cells (Fig. 2A). Neither terminally differentiated Hopx+/Lgfbp2+ nor Hopx+/Lgfbp2- cells expressed Ace2, indicating that AT1 cells are probably resistant to SARS-CoV-2 infection. Although it is generally accepted that SARS-CoV preferentially infects AT2 cells\(^35\), SARS-CoV-2 can infect both AT1 and AT2 cells ex vivo\(^40\) and in macaques\(^37\). These data may indicate the presence of alternative yet unidentified pathways of SARS-CoV-2 entry into AT1 cells.

The major contributor to alveolar epithelial renewal is a subpopulation of AT2 cells, which serve as alveolar stem cells\(^1,11\). These cells, which express Axin2, are referred to as alveolar epithelial progenitors (AEPs)\(^11\). The bulk RNA-seq data for Axin2+ AEPs and Axin2- AT2 cells\(^14\) were reanalyzed starting from raw reads, and transcript abundances were estimated by kallisto\(^46\). The expression of Ace2, Tmprss2 and Dpp4 was 2.68-fold, 3.79-fold and 2.31-fold, respectively, higher in AEPs than in Axin2- AT2 cells (Fig. 2B). Additionally, the data on SARS-CoV-2 entry factor expression were extracted from another scRNA-seq dataset\(^35\). The percentage of Axin2- AT2 cells with Ace2 expression was higher than that of differentiated AT2 cells (4.20% and 2.23%, respectively), but the level of Ace2 expression was similar in Axin2+ and Axin2- cells (Fig. 2C). Interestingly, the number of cells expressing Tmprss2 was low. Nevertheless, both datasets indicate that alveolar stem cells (i.e., AEPs) could be even more sensitive to SARS-CoV-2 infection than differentiated AT2 cells.

The alveolar epithelium has a relatively low regeneration potential; therefore, distal airway stem cells mobilize after lung injury to occupy alveolar surfaces. Alveolar regeneration in humans after SARS-CoV-2 infection has not yet been described, but epithelial regeneration in small bronchioles has been demonstrated in a nonhuman primate model\(^36\). Numerous distinct populations of stem cell types have been reported to contribute to regeneration after injury\(^6\), but we found RNA-seq datasets for only two types of stem cells.
Figure 1. The proportions of cells expressing SARS-CoV-2 entry factors are underestimated in standard scRNA-seq datasets. (a) Expression of the SARS-CoV-2 entry factors Ace2, Tmprss2, Furin, Anpep and Dpp4 in mouse trachea datasets from Montoro et al.\(^2\): 3′ scRNA-seq dataset (left panel) and full-length scRNA dataset (right panel). For the 3′ scRNA-seq dataset, unique molecular identifier (UMI) counts were normalized to account for differences in coverage, multiplied by a scaling factor of 10,000 to generate transcripts per kilobase million (TPM)-like values, and then log transformed. TPM values from the full-length scRNA dataset were rescaled to sum to 10,000 and were log transformed. Gene expression estimates were summarized in accordance with the cell type labels provided in the original paper. The dot size indicates the proportion of cells among the respective cell type population with greater-than-zero expression of the respective SARS-CoV-2 entry factor, while the dot color indicates the average nonzero expression value. (b) Correlation between gene expression in the 3′ scRNA-seq dataset and the full-length scRNA dataset for basal cells, club cells and ciliated cells. The Ace2, Tmprss2, Furin, Anpep and Dpp4 expression levels are represented by colored dots. (c) Full-length scRNA-seq detects a substantially higher number of cells with greater-than-zero expression of genes, including SARS-CoV-2 entry factors, among basal cells, club cells and ciliated cells. The percentages of cells expressing Ace2, Tmprss2, Furin, Anpep and Dpp4 are represented by the colored dots.
Figure 2. Lung epithelial stem cells express SARS-CoV-2 entry factors. (a) Expression of Ace2, Tmprss2, Furin, Anpep and Dpp4 in the subpopulation of AT1 cells that maintain the ability to transdifferentiate into AT2 cells (Hopx+/Igfbp2− cells) and in terminally differentiated AT1 cells (Hopx+/Igfbp2+ cells). The results for the scRNA-seq dataset from mice at postnatal day 633 are presented. The dot size indicates the proportion of cells in the respective cell type with greater-than-zero expression of the respective SARS-CoV-2 entry factor, while the dot color indicates the average nonzero expression value. (b) Volcano plot showing elevated expression of Ace2, Tmprss2, Furin, Anpep and Dpp4 in AEPs compared to differentiated AT2 cells (bulk RNA-seq dataset11). Each dot represents one gene. The log2 (fold change) in the expression levels of Ace2, Tmprss2, Furin, Anpep and Dpp4 is indicated by the colored dots (red, differentially expressed genes (p.adjusted < 0.05); pink, non-differentially expressed genes). (c) Expression of SARS-CoV-2 entry factors in AEP AT2 cells expressing Axin2 in the dataset from Wang et al.33. (d) Volcano plots showing the expression levels of Ace2 and Tmprss2 in BASCs versus differentiated AT2 cells (left panel) or differentiated club cells (right panel). The log2 (fold change) in the expression levels of differentially expressed genes (p.adjusted < 0.05) is shown in red; of non-differentially expressed genes, in pink. (e) Expression of SARS-CoV-2 entry factors in BASCs, AT2 cells, secretory club cells and ciliated cells. The proportion of Ace2-expressing cells in the BASC population was higher than that in the alveolar AT2 cell population but lower than that in the bronchiolar club cell population. (f) t-SNE plot of 1952 scRNA-seq profiles from the scRNA-seq dataset40, colored by cluster assignment. Cluster 6 contains H2-K1high club-like stem cells (Supplementary Information Figs. 2c and 2d). (g) SARS-CoV-2 entry factors are expressed in a large proportion of H2-K1high club-like stem cells as well as in other clusters of club cells and AT2 cells.
BASCs are activated and respond distinctly to different lung injuries; they can differentiate into multiple cell lineages, including club cells and ciliated cells of the terminal bronchioles and AT1 and AT2 cells of the alveoli. Our reanalysis of a bulk RNA-seq dataset indicated that BASCs express elevated levels of Ace2 and *Inpp5r*2 compared to AT2 cells (Fig. 2d, left panel). In contrast, club cells expressed higher levels of these two entry factors than BASCs (Fig. 2d, right panel). We also reanalyzed a scRNA-seq dataset to achieve better separation of the BASC cluster from other cell clusters in the t-SNE plot and refine the cluster labels (Supplementary Information Fig. 2a). BASCs coexpressed markers of AT2 cells (Sftpc) and club cells (*Scgb1a1*) (Supplementary Information Fig. 2b). Both the number of Ace2-expressing cells and the expression levels of Ace2 were higher in Bacs than in AT2 cells but lower in BASCs than in club cells (Fig. 2e). Interestingly, the proportion of Ace2-expressing ciliated cells was roughly equal to that of BASCs, but the Ace2 expression level in ciliated cells was substantially lower.

(ii) Recently, a rare population of H2-K1*high* club cell-like stem cells has been described. These cells, which differentiate into AT1 and AT2 cells following bleomycin-induced lung injury, were identified in scRNA-seq datasets of murine distal airways. Since the cell type annotation for H2-K1*high* cells was unavailable, we reanalyzed this dataset (Fig. 2f). Club-like cells identified by the expression of progenitor genes (Supplementary Information Fig. 2c) and secretory cell marker genes (Supplementary Information Fig. 2d) were reclustered into six subpopulations, including a H2-K1*high* cell population. A relatively large proportion of H2-K1*high* cells expressed Ace2 (~37.93%) and other viral entry factors, indicating that this type of stem cell is also sensitive to SARS-CoV-2 infection (Fig. 2g).

**Discussion**

SARS-CoV-2 can infect lung cells and induce severe respiratory failure. Recent scRNA-seq studies demonstrated that only a minor fraction of lower respiratory tract epithelial cells expressed SARS-CoV-2 entry factors. If the proportion of virus-sensitive cells is small, this may prevent or restrict infection of the lower respiratory tract. We suggested that the proportion of epithelial cells expressing SARS-CoV-2 entry factors could be underestimated due to the dropout effect. To roughly estimate the influence of the dropout effect, we compared two scRNA-seq datasets obtained using 3′ scRNA-seq and full-length scRNA-seq and found that the number of cells expressing Ace2 and other SARS-CoV-2 entry factors was indeed substantially lower in the 3′ scRNA-seq dataset. For example, the proportions of Ace2*high* basal cells were 0.60% and 9.38% in the 3′ scRNA-seq and full-length scRNA-seq datasets, respectively, indicating a strong influence of the dropout effect on the estimation of the number of Ace2-expressing cells. Hence, the fraction of cells that are potentially sensitive to the SARS-CoV-2 infection might be significantly higher than that described in the majority of published reports.

After episodes of infection or different injuries, the respiratory epithelium can be rapidly reconstituted by multipotent stem cells. Here, we analyzed published RNA-seq datasets to ascertain whether different lung epithelial stem cell types express SARS-CoV-2 entry factors. Because of the limited data on lung stem cells in humans, we conducted this study on data obtained from mice. Respiratory epithelium and epithelial stem cells are slightly different in mouse and human, but the patterns of SARS-CoV-2 entry factor expression in different epithelial cells of mouse and human airways are highly similar (Supplementary Information Fig. 1), indicating that mouse datasets can be used for our study in the absence of valid human datasets.

We found that epithelial stem cells (basal cells, AEPs, BASCs and H2-K1*high* cells) express Ace2 and other SARS-CoV-2 entry factors, making these cells probable targets of SARS-CoV-2 infection. The expression of these factors in different stem cells was relatively low, but, for example, among cells expressing markers that are specific to the gas-exchanging alveoli, AEPs exhibited higher expression of SARS-CoV-2 entry factors than differentiated AT1 and AT2 cells. These results are in agreement with the observations that cell differentiation is accompanied by depletion of the ACE2 protein.

Multipotent stem cells can reconstitute lung epithelium after episodes of infection or other injuries, and it was demonstrated that stem cells could proliferate in COVID-19 patients. However, the expression of SARS-CoV-2 entry factors makes them potentially infectable by SARS-CoV-2, which may in turn result in a decreased capacity for lung epithelial regeneration and potentially complicate recovery from the disease.

**Methods**

All RNA-seq datasets used in this study were retrieved from public data repositories. Detailed information including accession codes is available in Supplementary Information Table 1.

**Bulk RNA-seq analysis.** The data from the GSE97055 and GSE129440 datasets were reprocessed from raw reads and raw gene counts, respectively. Raw pair-end reads were quality and adapter trimmed with BBDDuk (minlen = 31 qtrim = r trimq = 20 ktrim = r k = 25 mink = 11 hdist = 1) from the BBTools suite, and FastQC was used for quality control. Then the reads were pseudoaligned to the mouse transcriptome [obtained from GRCm38 primary genome assembly and GENCODE gene annotation version M24 (https://www.gencodegenes.org/mouse/release_M24.html)] using kallisto with default parameters. TPM (transcripts per million) values provided by kallisto were imported and summarized into a gene-level matrix using the tximport R package. Differential gene expression analysis was carried out using the DESeq2 R package. Genes with Benjamini-Hochberg adjusted p-values < 0.05 were declared differentially expressed. Volcano plots were generated using the ggplot2 R package.

**Single cell RNA-seq (scRNA-seq) analysis.** For the GSE103354 dataset containing both 3′-droplet-based and full-length plate-based scRNA-seq data UMI (unique molecular identifier) counts and TPM values were available. 3′ scRNA-seq data were normalized using the NormalizeData function from the Seurat R
were identified by the absolute expression of marker genes: \texttt{Hopx} for \texttt{bors function}, \texttt{k.param = 20}, finding clusters (\texttt{FindClusters function}, \texttt{resolution = 0.4}) and subsequent significant PCs (identified by inspecting the elbow plot) were used for constructing the \texttt{SNN graph} (\texttt{FindNeighbors function with default parameters}), linear dimensional reduction (PCA) was conducted based on were retained for further analysis. After the data were normalized by the means of the \texttt{Seurat package} (\texttt{Nor - malizeData function with default parameters}), \texttt{mean TPM}) and percent of gene-expressing cells for clusters of basal, club, and ciliated cells based on cell labe-

clusters were pooled into one large cluster of secretory club-like cells based on the expression of known club cell markers. Igfbp2- cell subpopulation from the bulk AT1 cells, as well as \texttt{Axin2+} cells from AT1 cells, subclusters of interest were identified by the absolute expression of marker genes: \texttt{Hopx} and \texttt{Igfbp2} for AT1 cells and \texttt{Axin2 and Sftpc} for AT2 cells.

For the GSE129937 dataset raw UMI counts were available. Low-quality cells were filtered out with a mitochondrial fraction threshold of 10% and a number of detected genes threshold of 1000; 3278 high-quality cells were retained for further analysis. After the data were normalized by the means of the Seurat package (NormalizedData function with default parameters), linear dimensional reduction (PCA) was conducted based on 2000 HVGs (identified by the \texttt{FindVariableFeatures function with default parameters}). Scores from the first 8 significant PCs (identified by inspecting the elbow plot) were used for constructing the \texttt{SNN graph} (\texttt{FindNeighbors function}, \texttt{k.param = 20}), finding clusters (\texttt{FindClusters function}, \texttt{resolution = 0.4}) and subsequent \texttt{t-SNE visualization} (\texttt{RunTSNE function}, \texttt{perplexity = 50}). The described analysis resulted in the identification of 10 clusters of single cells. Based on cluster marker genes obtained by the \texttt{FindAllMarkers function (only.pos = TRUE, test.use = \texttt{“MAST”; requires MAST R package}) cell type identities were assigned to clusters. Cells from 5 smaller clusters were pooled into one large cluster of secretory club-like cells based on the expression of known club cell markers. The pooled dataset of club-like cells were then analyzed independently. To identify the \texttt{H2-K1\textsuperscript{high} cell population with a high progenitor gene signature supervised clustering using club cells HVGs and the list of Sox9-associated progenitor genes} was carried out. A Sox9-based progenitor gene list combined with HVGs was used as a feature input for PCA. The first 10 PCs declared as significant by inspecting the elbow plot were then used to find clusters (\texttt{FindClusters function}, \texttt{k.param = 20, resolution = 0.5}) and visualize them by the \texttt{t-SNE algorithm} (\texttt{RunTSNE function, perplexity = 50}). \texttt{H2-K1\textsuperscript{high}} specific genes from Kathiriya et al.\textsuperscript{46} were used to locate the cluster of interest.

For all scRNA-seq datasets the average expression values (log of mean TPM) of SARS-CoV-2 entry factors and percent of cells expressing SARS-CoV-2 entry factors were calculated for every cluster of interest using the \texttt{DotPlot function from Seurat.}

Code availability
R notebooks are available at https://github.com/kirushka/covid19_LSC.

Received: 15 July 2020; Accepted: 1 October 2020
Published online: 20 October 2020

References
1. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi. org/10.1038/s41586-020-2012-7 (2020).
2. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China. 2019. N. Engl. J. Med. 382, 727–733. https://doi.org/10.1056/NEJMoa2001017 (2020).
3. Mason, R. J. Pathogenesis of COVID-19 from a cell biology perspective. Eur. Respir. J. 55, 2000607. https://doi.org/10.1183/13993 003.00607-2020 (2020).
4. Hogan, B. L. M. et al. Repair and regeneration of the respiratory system: Complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell 15, 123–138. https://doi.org/10.1016/j.stem.2014.07.012 (2014).
5. Chen, J. Origin and regulation of a lung repair kit. Nat. Cell Biol. 19, 885–886. https://doi.org/10.1038/ncb3585 (2017).
6. Basl, M. C. et al. The cellular and physiological basis for lung repair and regeneration: past, present, and future. Cell Stem Cell 26, 482–502. https://doi.org/10.1016/j.stem.2020.03.009 (2020).
7. Rock, J. R. et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc. Natl. Acad. Sci. U.S.A. 106, 12771–12775. https://doi.org/10.1073/pnas.0906850106 (2009).
8. Pardo-Saganta, A. et al. Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell sub- populations. Cell Stem Cell 16, 184–197. https://doi.org/10.1016/j.stem.2015.01.002 (2015).
Acknowledgments
We are grateful to A.A. Penin, D.M. Potashnikova and A.A. Saidova for their helpful discussion and assistance. This work was supported by the Russian Science Foundation grant 18-14-00195 to E.V.S. Computational resources of the Makarich HPC cluster were provided by the Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University.

Author contributions
A.A.V., A.A.Z., A.S.K. conducted the experiments, analyzed the data and wrote the paper; Y.S.V. analyzed the data and wrote the paper; E.V.S. conceived and designed the study, analyzed the data and wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-74598-5.

Correspondence and requests for materials should be addressed to Y.S.V. or E.V.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020