



OPEN

## Mitochondrial gene expression in single cells shape pancreatic beta cells' sub-populations and explain variation in insulin pathway

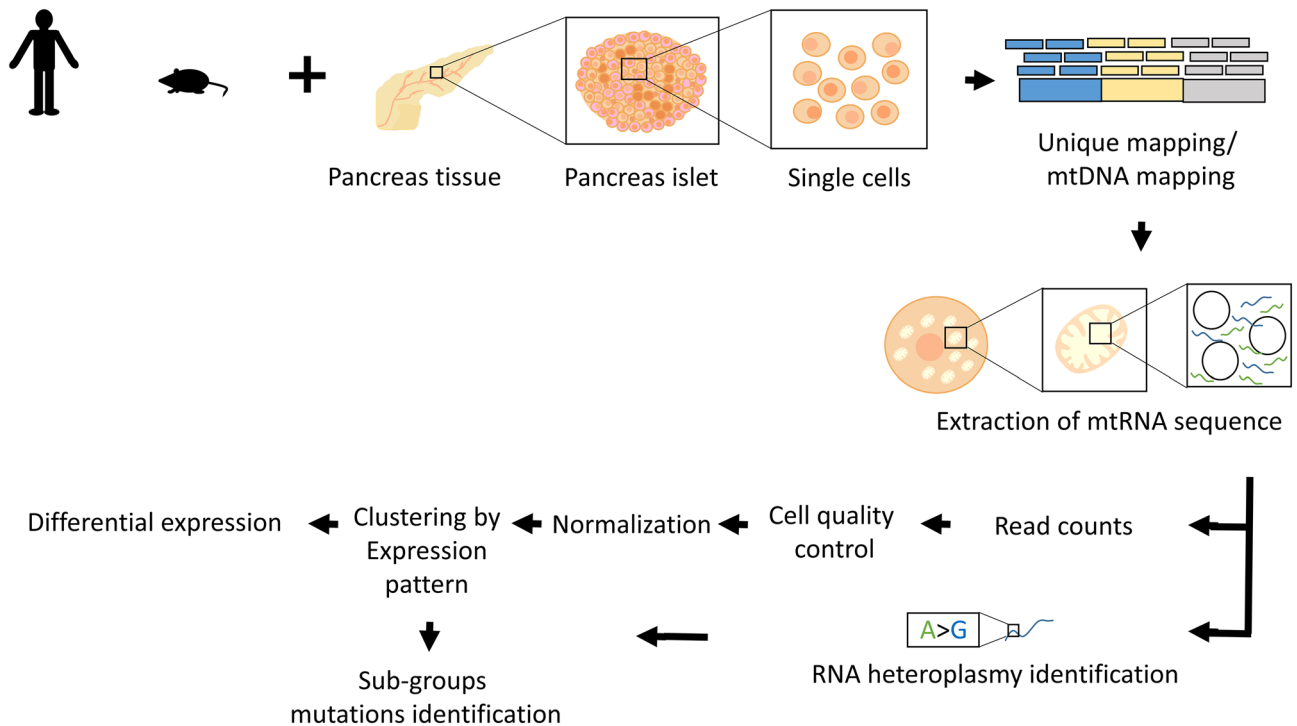
H. Medini, T. Cohen & D. Mishmar✉

Mitochondrial gene expression is pivotal to cell metabolism. Nevertheless, it is unknown whether it diverges within a given cell type. Here, we analysed single-cell RNA-seq experiments from human pancreatic alpha (N = 3471) and beta cells (N = 1989), as well as mouse beta cells (N = 1094). Cluster analysis revealed two distinct human beta cells populations, which diverged by mitochondrial (mtDNA) and nuclear DNA (nDNA)-encoded oxidative phosphorylation (OXPHOS) gene expression in healthy and diabetic individuals, and in newborn but not in adult mice. Insulin gene expression was elevated in beta cells with higher mtDNA gene expression in humans and in young mice. Such human beta cell populations also diverged in mitochondrial RNA mutational repertoire, and in their selective signature, thus implying the existence of two previously overlooked distinct and conserved beta cell populations. While applying our approach to human alpha cells, two sub-populations of cells were identified which diverged in mtDNA gene expression, yet these cellular populations did not consistently diverge in nDNA OXPHOS genes expression, nor did they correlate with the expression of glucagon, the hallmark of alpha cells. Thus, pancreatic beta cells within an individual are divided into distinct groups with unique metabolic-mitochondrial signature.

Mitochondrial metabolism is pivotal for the function of all cells, yet it is especially critical for energy demanding tissues, such as brain, muscle and pancreatic beta cells. The hallmark of pancreatic beta cells' activity is insulin secretion, which is compromised in type 1 diabetes, and to a lesser extent in type 2 diabetes mellitus (T2DM)<sup>1</sup>. This process is triggered by increase in concentrations of cytosolic Ca<sup>2+</sup> in response to depolarization of the plasma membrane, and require ATP produced by the mitochondria of beta cells, via the oxidative phosphorylation system (OXPHOS)<sup>2,3</sup>. OXPHOS employs five multi-subunit protein complexes harbouring 13 mitochondrial DNA (mtDNA)-encoded subunits and ~80 nuclear DNA (nDNA)-encoded proteins<sup>4</sup>, which physically interact and co-evolve<sup>5,6</sup>. Such bi-genomic system is the result of genes' transfer from the genome of the once free-living mitochondrial ancestor to the host nucleus<sup>7</sup>. This transfer of genes with prokaryotic heritage into the nuclear genome, led to their individual transcription, in different from the few remaining mtDNA-encoded genes which are transcribed in bacterial-like strand-specific polycistrons<sup>8,9</sup>. Despite the apparent difference in transcriptional regulation, mtDNA- and nDNA-encoded OXPHOS genes are co-expressed across many tissues<sup>10,11</sup>. This suggests that regulation of mitochondrial genes' expression likely adapted to its host, in order to coordinate the activities of both the mitochondrial and nuclear genomes. Deficient transcription and translation in mitochondria have been shown to impair stimulus of beta cells insulin secretion<sup>12</sup>. Specifically, insulin secretion was severely impaired upon conditional knockout of transcription factor A (TFAM) and transcription factor B2 (TFB2M) in mouse beta cells<sup>13,14</sup>. In addition, TFAM expression was reduced with disruption of transcription factor PDX1, which controls embryonic development of the pancreas and function of mature beta cells. Moreover, beta cells mitochondria generate metabolites that couple glucose sensing to exocytosis of insulin granule<sup>15</sup>, thus underlining the importance of mitochondrial function to the fundamental activity of pancreatic beta cells. Specifically, these findings strongly suggest that mtDNA regulation is essential for insulin secretion.

It has been previously suggested, that pancreatic beta cells are heterogeneous in terms of gene expression, cell surface antigens<sup>16</sup>, metabolic capacity<sup>17</sup>, and rates of insulin synthesis<sup>18</sup>. Although it has been shown that ATP-stimulated insulin secretion correlate with mitochondrial signalling<sup>15,19</sup> and relies on active mtDNA regulation<sup>13</sup>, it is yet unclear whether beta cells are homogenous in regulation of mitochondrial gene expression, and whether

Department of Life Sciences, Ben-Gurion University of the Negev, 8410501 Beer Sheva, Israel. ✉email: dmishmar@bgu.ac.il



**Figure 1.** Workflow of scRNA-seq analysis. Hypothesis free scRNA-seq clustering according to mtDNA gene expression. Fastq-files were mapped against the entire genome (GRCh38 for human cells and GRCm38 for mouse cells). After mapping, we calculated read counts, followed by data quality assessment, clustering and differential expression analyses. The RNA mutational heterogeneity data was used to identify mutations that characterize each of the identified cell groups per individual, in the Fluidigm C1 platform (see Methods).

such putative variability affects beta cell activity. To address this question, we conducted hypothesis-free analyses of mitochondrial gene expression in four publicly available single cell RNA-seq (scRNA-seq) experimental datasets of beta and alpha cells from T2DM and healthy human donors, as well as in four scRNA-seq datasets from mouse.

## Results

**Analysis of human scRNA-seq data reveals mtDNA gene expression divergence between alpha and beta pancreatic islet cells.** As a first step, scRNA-seq from human pancreatic beta and alpha cells were analysed in three publicly available datasets<sup>20–22</sup> (Fig. 1, Table S1). The single-cell human transcriptomic Datasets I (InDrops sequencing protocol) contained ~ 10,000 human pancreatic cells from four donors isolated from three non-diabetic individuals (ND) and one T2DM patient. Dataset II (Fluidigm C1 sequencing protocol) contained a total of 1492 pancreatic cells from twelve human ND and six T2DM donors, and Dataset III (Fluidigm C1 as in Dataset II) contained 638 pancreatic cells from five ND and three T2DM patients. After filtering of Dataset I, while applying quality control measures (taking into account zero inflated reads, a minimum of genes' number per cell, maximum representation of rRNA transcripts, propensity for doublet cells—see details in Methods), we were left with a total of 2776 cells (1827 alpha cells and 949 beta cells) with ~ 2500 informative genes on average per cell. For Datasets II and III, cells with less than 3000 genes were excluded (due to the different sequencing technology as compared to Dataset I, and the higher sequencing depth of the Fluidigm C1 platform). This resulted in 1396 cells from Dataset II (928 alpha cells and 468 beta cells), and 491 cells from Dataset III (239 alpha cells and 252 beta cells) with ~ 5000 and ~ 6000 informative genes on average per cell for subsequent gene expression analysis, respectively (Table S1). Cells that were called either alpha or beta cells expressed their characteristic transcript, namely either insulin (INS; beta-cells) or glucagon (GCG; alpha-cells)<sup>22,23</sup> (Fig S1, S2).

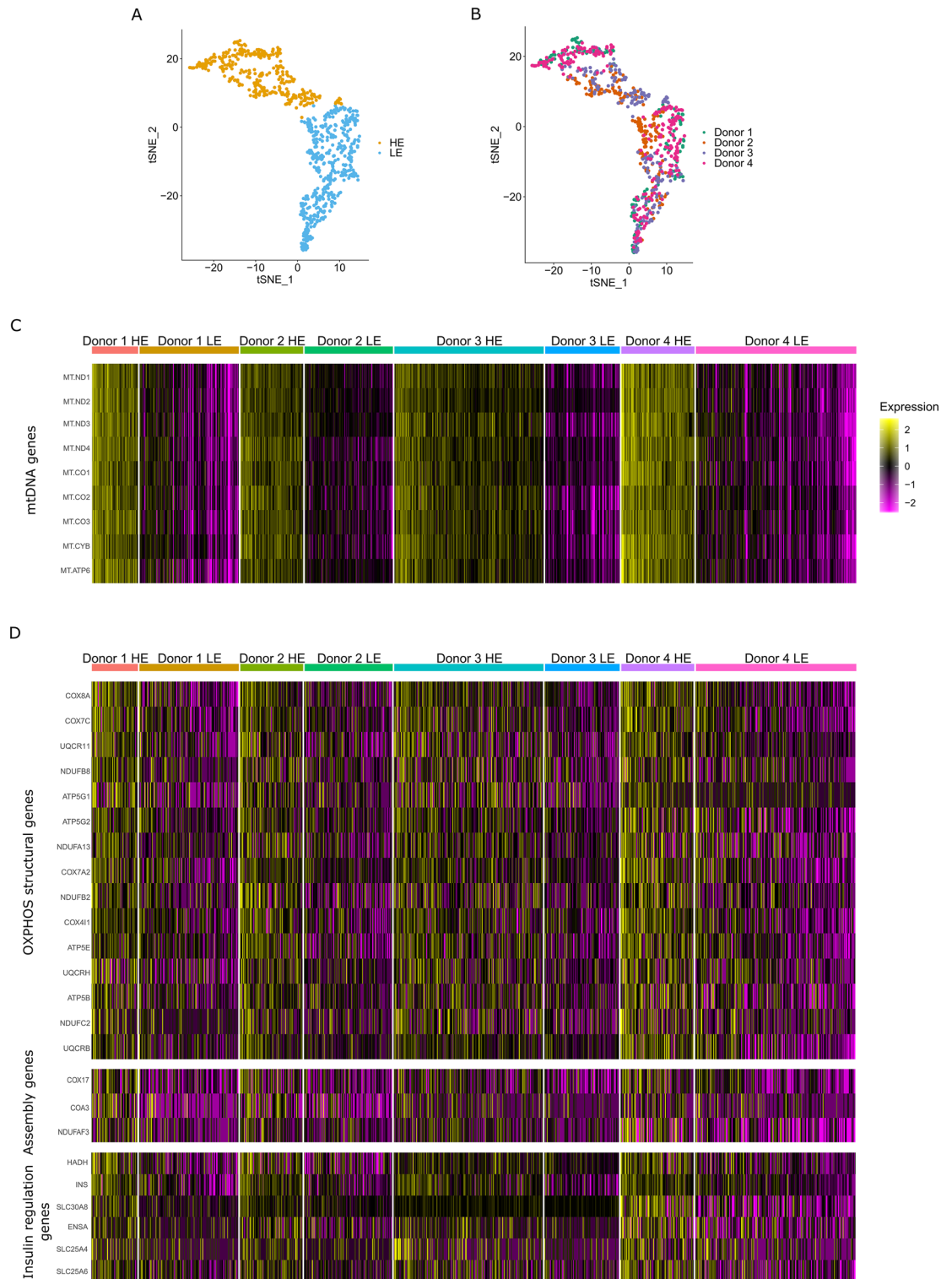
Comparison of mtDNA gene expression between alpha and beta cells, in healthy and T2DM donors, revealed significantly higher mtDNA transcript levels in beta cells both in healthy and in T2DM patients as compared to alpha cells in all studied datasets (Table S2) (Fig S3–5). These findings are consistent with known metabolic functional differences between alpha and beta cells in humans<sup>24</sup>. Notably, due to the very small sample size we could not directly compare healthy and diabetic individuals, as well as assess heterogeneity of the donors in terms of ethnicities and gender (Table S2). The consistency of these findings with previously-published studies encouraged us to continue our analysis further into investigating the subpopulations within each of the tested cell types.

**Human beta cells diverge according to expression of mtDNA and nDNA-encoded OXPHOS genes.** The function of pancreatic beta cells relies on mitochondrial activity. Nevertheless, it is unclear

whether beta cells are homogenous in mitochondrial regulation. As a first step to address this question, we performed hypothesis-free cluster analysis of mtDNA genes' expression in beta cells from the four donors in Dataset I using Seurat<sup>25</sup>. This analysis identified two clearly distinct beta cells' clusters across all tested donors, which differed in mtDNA gene expression per donor (i.e. two groups with either high or low gene expression, designated HE and LE, respectively) (Fig. 2A–C). The beta cell subgroups were consistently identified even when analysing all the individuals per dataset grouped together (Fig. 2B). Accordingly, the latter analysis revealed that ~84% of the cells in Dataset I remained in their original subgroups, regardless of the donors, thus further supporting the robustness of these clusters. We next analyzed Datasets II and III to determine the robustness of these subgroups. To avoid sample size issues, we controlled for sample sizes per donor, and limited our analysis to donors with a minimum of high quality RNA-seq from at least 40 beta cells (Table S1). Despite the small cell number per donor, our findings revealed sharp division into two cell clusters which diverged in the levels of mtDNA gene expression (Fig S6). When analysing all the individuals per dataset together such subgroups were consistently identified as in Dataset I (Fig S6), namely ~86% and 87% of the cells in Datasets II and III, respectively, remained in their original subgroups, regardless of donors. This supported the robustness of these cell clusters, which were identified regardless of the sequencing platform used, and health status of the tested individuals (Fig S7). It is worth noting, that this result did not differ between the sequence mapping methods used, i.e. unique or default mapping (Fig S8).

Since mitochondrial activities are coordinated between the mtDNA and the nucleus<sup>10</sup>, we asked whether the expression of nuclear DNA-encoded (nDNA) genes associates with our observed beta cells populations. As a first step to address this question we analysed genes that consistently differentially expressed between the subgroups across all four donors in a selected set of ~300 nDNA-encoded proteins which are translated in the cytoplasm and are imported into the mitochondria<sup>10,26</sup>. This set of genes included all known factors that regulate mtDNA replication, transcription, translation and RNA stability, as well as assembly factors and structural subunits of the mitochondrial OXPHOS system<sup>27</sup>. Our analysis of Dataset I revealed that the expression of OXPHOS structural genes (complexes I, III, IV and V) consistently diverged between the LE and HE beta cell populations across all four donors (Fig. 2B, Table S3). Notably, although to a lesser extent, certain assembly factors of OXPHOS complexes I and III also correlated with these cell populations. To identify differentially expressed genes across individuals in Datasets II and III, we applied the same analysis to the six individuals available from these datasets. The combined analysis of the total samples of beta cells per dataset, revealed expression divergence of OXPHOS structural genes between the HE and LE cellular sub-groups (Fig S9, Table S3), including genes that were consistent between the three tested datasets. Therefore, our results indicate the discovery of novel sub-populations of human pancreatic beta cells that diverge in mito-nuclear OXPHOS gene expression. Finally, we noticed that previous studies used a certain mtDNA read percentage threshold (~10%) to avoid cells with apparently lower quality<sup>28</sup>. Although using such a threshold notably reduced the sample size of analysed cells per database (nearly tenfold in some cases), the HE and LE cellular groups prevailed, while retaining significant differences in both mtDNA and nDNA-encoded OXPHOS gene expression (Fig S10, Table S3). Notably, using higher mtDNA read threshold, calculated according to the median mtDNA read fraction per dataset (while removing cells with more than 2 SD higher mtDNA read counts), did not change the number of analysed genes, yet increased the numbers of analysed cells. Since this approach did not compromise cell quality and avoid usage of small sample sizes we applied it to the rest of our analyses.

**The two beta cell sub-populations in humans diverge in Insulin gene expression.** Next, we asked whether our identified beta cell subpopulations associate with the expression of other, additional genes across the human genome. To test for such, we extended our analysis to the entire human transcriptome. Given the set of differentially expressed nDNA genes from the combined analysis (beta cells from all four donors of Dataset I; Table S3) we applied an enrichment analysis to explore which biological processes (GO terms) differentially expressed in the two beta cell subgroups (Table S3). As expected, the analysis revealed that ATP metabolic process and OXPHOS were in the top-ten of the genes that were upregulated in the HE sub-group of cells (Table S3). We noticed, that the full list of significant processes also included genes involved in insulin regulation and secretion. Strikingly, we found that the cell cluster with higher mtDNA gene expression showed significantly higher expression of *INS*, encoding the insulin transcript ( $p < 1 \times 10^{-50}$ , Dataset I; FDR correction). To assess whether the differential expression of the insulin regulatory pathway is more prominent than other pathways, we assessed differential expression of selected gene pathways between the HE and LE groups of cells: regulation of insulin secretion, cell proliferation, glycolysis and cell cycle. This analysis revealed, that the HE cells' group showed significantly high expression of genes involved in regulation of insulin secretion, including the following: Firstly, *SLC30A8*, encoding a zinc-efflux transporter (zinc transporter 8 (ZnT8)) which mediates uptake of zinc into secretory granules ( $p < 0.05$ , Dataset I, FDR correction)<sup>29</sup>; secondly, *SLC25A4* and *SLC25A6* which translocate ADP from the cytoplasm into the mitochondrial matrix and ATP from the mitochondrial matrix into the cytoplasm<sup>30</sup>; third, *ENSA* which encodes an alpha-endosulfine, a regulator of the beta-cell K(ATP) channels ( $p < 0.05$ , Dataset I, FDR correction)<sup>31</sup>, and *HADH* gene, a negative regulator of Insulin secretion<sup>32</sup> (Fig. 2D). Finally, *PTPRN*, which participates in the beta cells proliferation pathway and normal accumulation of secretory vesicles ( $p < 1 \times 10^{-8}$ , Dataset I; FDR correction)<sup>33</sup> was also upregulated in the HE subgroup. Notably, *PPP1R15A*, an unfolded protein response (UPR) gene, that was previously associated with low *INS* gene expression levels in mouse beta cells<sup>34,35</sup>, was upregulated in the beta cells group with lower mtDNA gene expression (LE) ( $p < 0.0032$ , Dataset I; FDR correction). Additionally, the expression of *INS* was also consistently higher in the HE subgroup of beta cells in Datasets II and III ( $p < 1 \times 10^{-5}$ , Dataset II, FDR correction;  $p < 0.005$ , Dataset III, FDR correction), thus further attesting for the robustness of this result. It is worth noting, that while examining additional mito-nuclear genes we found that the expression of *MEF2D*—a transcription factor that was shown



**Figure 2.** Human mtDNA gene expression analysis revealed two distinct beta cell clusters with either high or low mtDNA gene expression (designated HE and LE, respectively). (A,B) tSNE distribution of beta cells from the four human donors (Dataset I) showing two subgroups of beta cells with high and low mtDNA gene expression. (A) Subgroups of cells with either high (HE) or low (LE) mtDNA gene expression (yellow and blue dots, respectively); (B) cells are divided into HE and LE according to mtDNA gene expression, regardless of donor identity (donor’s cells are colour coded as indicated). (C,D) Heatmap showing the significant differentially expressed genes per beta cell cluster, per individual (after FDR correction). (C) mtDNA-encoded transcripts, (D) Upper panel—OXPHOS structural genes; Middle panel—OXPHOS assembly genes; lower panel—genes involved in insulin regulation. Colour codes: purple—low expression, yellow—high expression.

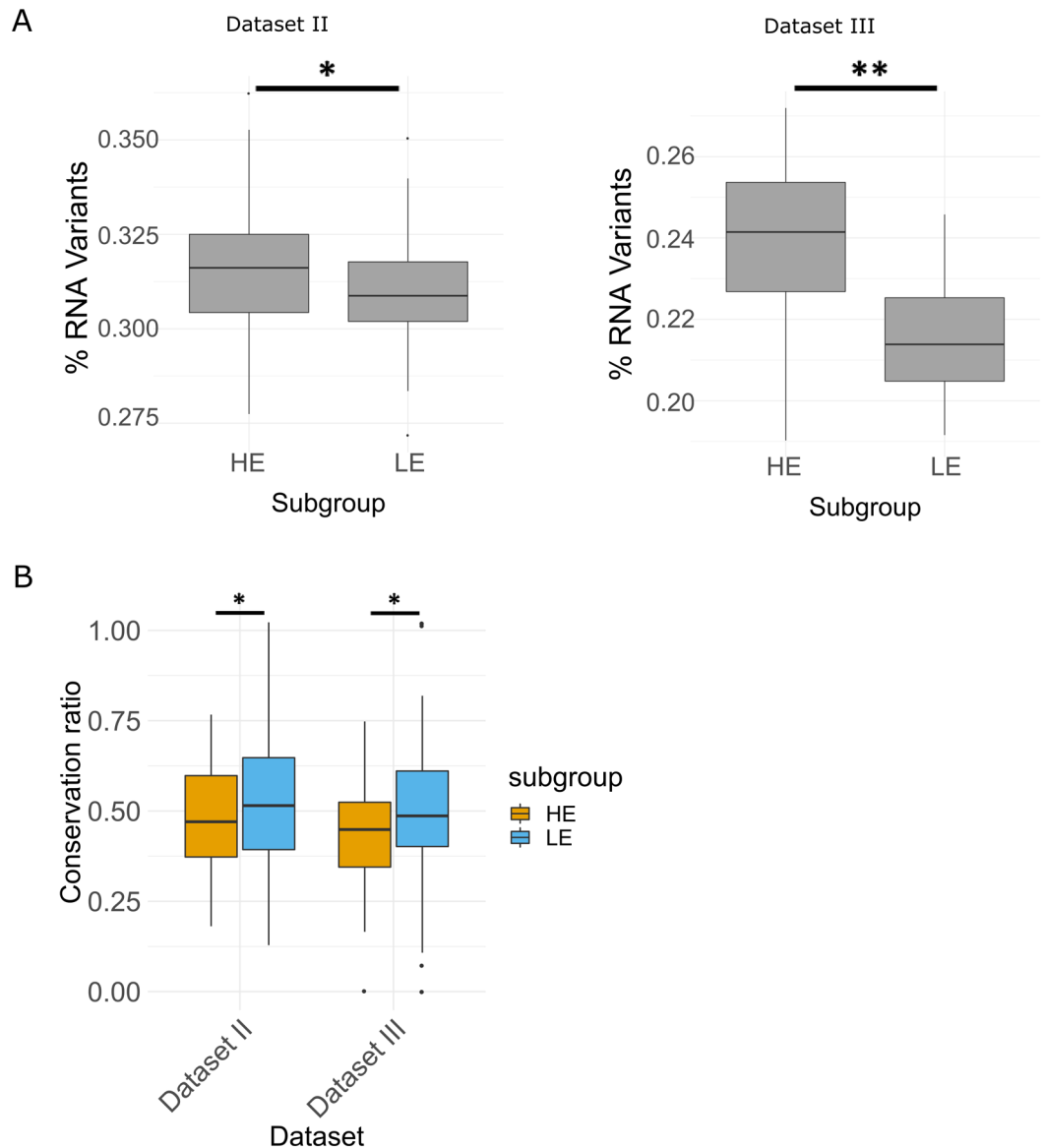


to regulate both nDNA and mtDNA gene expression<sup>36</sup>, was higher in the HE subgroup ( $p < 1 \times 10^{-6}$ , Datasets II, FDR correction;  $p < 1 \times 10^{-11}$ , Dataset III, FDR correction), thus suggesting an attractive candidate regulator, which explains differences between the HE and LE subgroups. Finally, we analyzed a fourth dataset of single beta cells RNA-seq, generated from four individuals using a different platform (Cell-seq2)<sup>37</sup> with an average of ~ 5600 informative genes per cell (designated here as human Dataset IV). Analysis of cells from three individuals having sufficient high quality beta cells RNA-seq data (N = 95, N = 82, N = 143 cells, Table S1) revealed, again, the LE and HE groups of cells which differed in mtDNA gene expression per individual. In similar to the above-mentioned human datasets I-III, analysis of all individuals together revealed that 99% of the cells remained in their original subgroups (LE and HE), regardless of tested individuals. These cell clusters positively correlated with OXPHOS genes expression and with the expression of insulin (Fig S11–S12, Table S3). These results not only further fortify our findings but underlines the robustness of these findings to the different sequencing techniques. Taken together, our mito-nuclear co-expression analysis strengthens the interpretation that pancreatic beta cells are divided into sub-populations which diverge in mitochondrial gene expression. This divergence is not only limited to mitochondrial activities, but also associates with the expression and regulation of insulin—the hallmark of beta cells' function. To our knowledge, these results serve as the first demonstration of mitochondrial regulatory involvement in physiologically relevant variability of beta cells activity.

Beta cells heterogeneity was previously mentioned in context of specific antigens and their expression<sup>38</sup>; although the identified subgroups of cells in Dorrell et al. (2016) are not apparently associated with mitochondrial function, we assessed whether our identified groups of cells correlate with this published sub division of human beta cells. Our analysis did not support correlation between the expression of the antigens identified by Dorrell et al. with our identified sub-groups of beta cells (Fig S13).

**RNA mutational repertoire is elevated along with mtDNA genes' expression.** The divergence of beta cells according to the expression of both mtDNA and certain nDNA genes suggests that human pancreatic beta cells are divided into two populations with distinct mitochondrial profiles. We therefore asked whether these two sub-populations of cells also diverge in pattern of mtDNA mutations. Notably, unlike the nuclear genome, mitochondrial RNA (mt-RNA) sequence heterogeneity could stem from mtDNA sequence variation (heteroplasmy), RNA sequence heterogeneity (due to RNA polymerase errors) and RNA modifications, as we recently discovered<sup>39–41</sup>. To identify mt-RNA mutations with high confidence, we determined RNA heterogenic mitochondrial mutations with a computational pipeline that utilized individual per-base sequence differences, while employing quality control measures to avoid sequencing errors. As mentioned above, due to low sequence coverage at the non-coding mtDNA region, we focused our analysis on the protein coding mtDNA sequences. Then, we verified that each tested individual had more than a 1000 mtDNA positions with high sequence coverage (> 400×). This requirement enabled analysing the sequences generated in Datasets II and III, but not in Datasets I and IV, which displayed lower per base coverage (Datasets I and IV contained on average ~ 100,000 and ~ 41,000 reads, respectively, for each analysed cell as compared to an average sequencing depth of  $0.95 \pm 0.46$  million reads and 34 million reads in Datasets II and III, respectively). While interrogating the repertoire and distribution of the RNA heterogenic mutations we found greater mitochondrial RNA (mt-RNA) mutational repertoire in the HE group as compared to the LE cell cluster in both Datasets II and III ( $p < 0.005$ , Dataset II;  $p < 1 \times 10^{-16}$ , Dataset III) (Fig. 3A, Table S4). Additionally, we noticed lower number of overlapping mutations as compared to unique mutations of the subgroups (Table S4). Next, we divided the mt-RNA mutations into candidate inherited mutations and 'others', according to the following logic: mutations shared by two cell types (i.e., alpha and beta cells) have likely been present prior to embryo differentiation, whereas this cannot be argued for other mutations; hence, these two groups of mutations (i.e., candidate inherited mutations, and others) should be regarded as either enriched or low in inherited mutations, respectively. As intuitively expected, the percentage of candidate inherited mutations in each beta cells group was found to be higher in the overlapping mutations between the groups as compared to the unique mutations in each group, per individual (Table S4), and the proportion of unique mutations was higher in the LE group as compared to the HE group in five out of six individuals. To better understand the functional potential of the mutations in each subgroup, we tested whether RNA heterogenic mutations occurred randomly throughout the mtDNA, per subject. Interestingly, the observed mutational conservation score was lower than expected by chance in both groups, although the LE group had significantly higher score as compared to the HE group in both datasets ( $p < 0.05$ ) (Fig. 3B) and a tendency towards higher score in all six tested individuals (Fig S14). Thus, the two beta cells sub-groups differ in mitochondrial mutational repertoire, and in the potential impact of such mutations, suggesting a stronger signature of negative selection acting on mt-RNA mutations in the HE beta cells group.

**Glucagon and OXPHOS genes do not consistently co-express in human pancreatic alpha cells subpopulations.** As a first step to assess the generality of the distinct beta cell sub-groups to other pancreatic cell types, we took advantage of scRNA-seq data of pancreatic alpha cells from the same four datasets, considering only individuals with more than 40 alpha cells each. These criteria left us with all donors from Dataset I, seven human donors in Dataset II, including the three individuals that had sufficient numbers of beta cells; two human donors in Dataset III, excluding one individual in our above-described beta cells analysis and four human individuals in Dataset IV. After applying the same approach used for beta cells analyses, although alpha cells could be divided into sub-groups according to mtDNA gene expression (Fig S15, Fig S16, Fig S17, Fig S18) they co-expressed with certain nDNA-encoded OXPHOS genes across two datasets out of four (Datasets I and II) (Fig S15, Fig S18, Fig S19), yet such subgroups did not display significant expression difference of glucagon (GCG) between the two subgroups (Fig S15, Fig S18, Fig S19). Notably, GO terms analysis revealed a weaker association with the OXPHOS and ATP metabolic processes as compared to beta cells (Table S5). In summary,



**Figure 3.** mtRNA mutation patterns display higher mutational repertoire and lower conservation score in the HE subgroup. (A) Box plot showing the comparison of the mutational repertoire in protein coding mtDNA genes between the two subgroups (LE and HE), per dataset (Datasets II and III). (B) Box plot demonstration of the distribution of evolutionary conservation ratios in mutations within the LE and HE groups, per analysed dataset.

although both pancreatic alpha and beta cells could be divided into subgroups according to mitochondrial gene expression, alpha cells display weaker association of nDNA-OXPHOS genes with such cellular population division, and do not correlate with their hallmark gene expression. Therefore, while considering human pancreatic islets, the association of mitochondrial gene expression with the hallmark of cellular gene expression is limited to beta cells.

### Mito-nuclear genes' expression and insulin define beta cell sub-groups in new-born, but not in adult mice.

We next asked whether the phenomenon of two distinct beta cell sub-groups in humans, which are divided according to mtDNA gene expression, is conserved in evolution. The available scRNA-seq mouse datasets originate from three studies of the C57Bl6 mouse strain (termed mouse Datasets I, II, III)<sup>22,42,43</sup>, with mouse Dataset I yielding 551 single beta cells<sup>22</sup>, 314 single beta cells in Dataset II from 3–7 month-old mice<sup>42</sup> and 387 beta cells from multiple postnatal time points in Dataset III collected from new-born mice (e.g., 84 cells collected from day 1, 87 cells from day 7, 88 cells from day 14, 68 cells from day 21, and 60 cells from day 28 postnatal)<sup>43</sup>. Similar to humans, mouse Dataset I was sequenced by the InDrops platform, whereas Datasets II and III were sequenced by Fluidigm C1. After quality control analysis (see Methods), 264 single beta cells remained for further analysis from mouse Dataset I, while considering ~1600 genes genome-wide on average

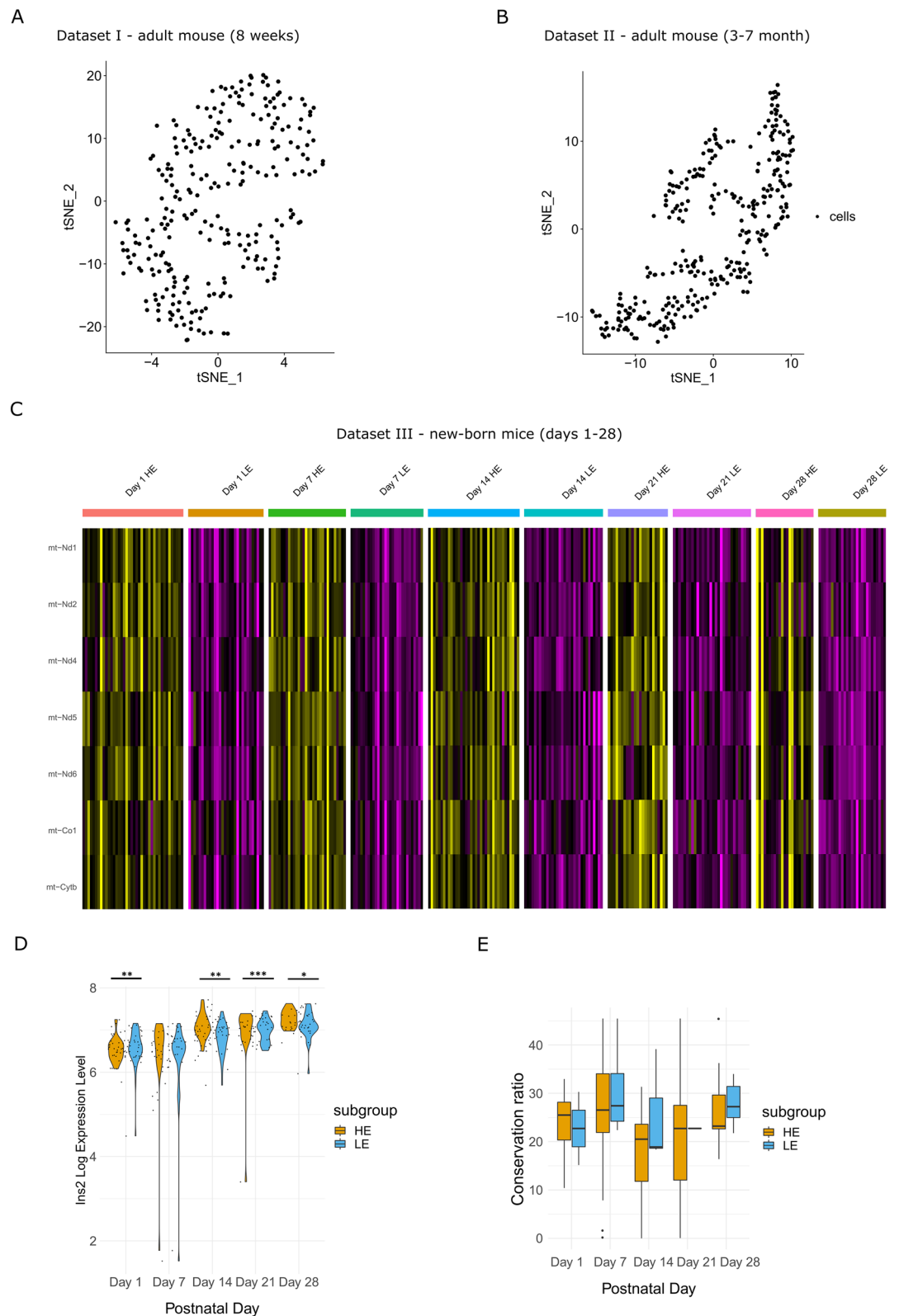
per cell. Filtering cells and genes in mouse Dataset II resulted in 309 single beta cells with an average of ~5700 informative genes per cell. In mouse Dataset III (new-born mice) a total of 304 beta cells remained, including 70, 62, 69, 53 and 50 single cells for mice from day 1, day 7, day 14, day 21 and day 28, respectively, with ~6500 informative genes on average, per cell. As in humans, the InDrops platform (Dataset I) enabled us analysing mtDNA protein-coding transcripts with >10 PolyA nucleotides (excluding Nd4l, Atp8 and Nd6; Table S6)<sup>44</sup>. To control for high similarity (99.9%) of the mouse mtDNA sequences overlapping the genes Nd3, Nd4l, Cox2, Cox3, Atp6, Atp8 with several nuclear mitochondrial mouse pseudogenes (NUMTS)<sup>45</sup>, we first limited our analysis to the seven remaining mtDNA protein-coding genes. This analysis revealed a single group of beta cells in adult mice (mouse Datasets I, II) (Fig. 4A,B). In contrast, Dataset III (new-born mice) displayed two subgroups of beta cells (Fig. 4C) in each of the available postnatal days, with one beta cells cluster showing higher mtDNA gene expression levels (in all tested mtDNA-encoded genes). Differential expression analysis of the orthologues nDNA-encoded OXPHOS genes in new-born mice revealed co-expression of certain structural genes, which differed among the postnatal days (Table S7). Specifically, Day 1 showed significantly high expression of certain structural and assembly genes in the LE subgroup, which were also upregulated in the human HE subgroup. Day 7, and more prominently cells from days 14–28, displayed significant overexpression of certain OXPHOS structural genes in the HE group as in humans, although certain structural and assembly genes that were markers of the HE group in humans were upregulated in the LE group of cells from these days. When we extended the analysis to the entire genome we found significantly higher expression of Ins2 at postnatal day 1 ( $p < 0.001$ ) in the LE group and higher expression of Ins1 gene ( $p < 1 \times 10^{-5}$ ) in the HE group. In contrast, this analysis revealed significantly higher expression of Ins2 in the HE subgroup as compared to the LE subgroup at postnatal days 14 ( $p < 0.001$ ), 21 ( $p < 1 \times 10^{-5}$ ) and 28 ( $p < 0.05$ ) (Fig. 4D), and with significantly higher expression of Ins1 at day 14 ( $p < 0.01$ ), but not in day 7 (Table S7). Finally, unlike our analysis in human beta cells, comparison of the mt-RNA mutational repertoire between the mtDNA gene expression of the beta cell clusters from the new-born mice (Dataset III) at postnatal days 14 and 28 revealed significantly higher mutational repertoire in the LE group (i.e., with the lower mtDNA gene expression) ( $p < 0.005$ , day 14;  $p < 1 \times 10^{-10}$ , day 28) (Fig S20), while day 1 and day 21 showed higher mutational repertoire in the HE subgroup as in humans. Notably, similar to human, the percent of the overlapping mutations was lower than the percent of the unique mutations in the subgroups (Table S4). Nevertheless, conservation analysis of these mutations revealed that the observed ratios of each group were insignificant and inconsistent among the postnatal days (Fig. 4E). It is worth noting that the results withstood a different mapping approaches, namely mapping solely against the mtDNA, which enabled including all protein-coding mtDNA genes in the analysis (see methods, Fig S21, Table S4, Fig S22).

In a previous study a cluster of beta cells with higher expression of mitochondrial genes has been identified in adult mice<sup>46</sup>. Interestingly, these mice (assigned here as mouse dataset IV) were fed either by regular or by high fat diet, thus allowing testing for the possible impact of such environmental condition on gene expression. When we analysed the available ~300 beta cells from mouse Dataset IV (Fig S21) we identified two beta cell clusters which differed in mtDNA gene expression, thus confirming the published results. Nevertheless, such division of the cells into sub-groups did not correlate with the expression of OXPHOS genes, neither did they correlate with the expression of Insulin gene. Notably, this lack of correlation did not change in mice fed with either type of diet. In summary, these findings revealed clustering of mito-nuclear gene expression within new-born mouse beta cells, suggesting that although mt-RNA gene expression divided beta cells into subpopulation in both human and young mice, other attributes of these sub-group of cells (such as mt-RNA mutational repertoire and conservation score) diverge.

## Discussion

Taken together, this work revealed mitochondrial gene expression clustering in human pancreatic beta cells. Such heterogeneity was reflected by two distinct sub-populations of cells which diverged by mtDNA gene expression, nDNA OXPHOS and Insulin gene expression, and in patterns of negative selection acting on mt-RNA mutational repertoire. Since all mtDNA protein-coding genes comprise essential subunits of the OXPHOS, such differences between the sub-groups of beta cells most likely reflect previously overlooked beta cell populations divergence in terms of mitochondrial regulation, and activity. This interpretation is consistent with the positive correlation that we found with insulin gene expression. This yields a testable hypothesis—it would be of interest to test whether our observed sub-populations of beta cells correlate with beta cells activity. Such analysis will enable correlating mitochondrial gene expression with the presence of so-called insulin secreting hubs<sup>17</sup> having high glucokinase:insulin ratios, and the presence of ‘extreme’ beta cells with elevated mRNA levels of insulin versus ‘non-extreme’ beta cells that were identified in mouse<sup>47</sup>. While human beta cells presented with a profound mitochondrial regulatory difference between two cellular sub-groups, pancreatic alpha cells did not. Specifically, although we observed an apparent sub-division into cells with different mtDNA gene expression, the expression correlation with nDNA OXPHOS genes was weaker, and the connection to the inherent function of the cell—glucagon expression, was not evident. This suggests that the mitochondrial subdivision of beta cells into subgroups is not common to all islet cell types. Furthermore, as insulin secretion has been clearly shown to rely on mitochondrial function, and alpha cells function rely more on anaerobic glycolysis<sup>12,48,49</sup>, it is plausible that heterogeneity in mitochondrial regulation within a given cell type relies on the centrality of mitochondrial function to the tested cell type. Therefore, there is great interest in assessing mitochondrial regulatory heterogeneity in other additional cell types and tissues.

While considering mitochondrial regulatory heterogeneity in mouse beta cells, new-born mice displayed sub groups of cells which, similar to humans, diverge in their mtDNA gene expression patterns and correlated with Insulin gene expression. However, the characteristics of the subgroups in terms of nuclear gene expression changes during the development of the neonates, as days 14–28 displayed a more similar expression pattern to



**Figure 4.** Pancreatic beta cells from newborn, but not adult mice, are divided into two clusters according to mtDNA genes' expression. **(A)** tSNE profile of mtDNA genes' expression pattern (protein-coding genes) in beta cells from 8 weeks old mice (mouse Dataset I). **(B)** tSNE of mtDNA genes' expression pattern (as in A) from beta cells from 3–7 month old mice (mouse Dataset II). **(C)** Gene expression heatmap of mtDNA gene expression in beta cells reveal cellular sub-groups in newborn mice (1, 7, 14, 21 and 28 days postnatal—mouse Dataset III). **(D)** Ins2 gene expression in beta cell sub-groups in each of the tested postnatal days (mouse Dataset III). **(E)** Box plot of the evolutionary conservation ratio of mtDNA mutations in beta cells from new-born mice, per postnatal day (mouse Dataset III). Significance: \* $p < 0.05$ , \*\* $p < 0.001$ , \*\*\* $p < 1 \times 10^{-5}$ .

human as compared to days 1 and 7. This can stem from the immature metabolic phenotype of the neonatal beta cells in mice<sup>50</sup>. In contrast, while considering mitochondrial gene expression, the adult mice (8 weeks and 3–7 month) displayed a more homogenous population of beta cells. The observed differences between human and adult mouse beta cells might stem from the islets architecture<sup>51</sup> and the difference in longevity of the beta cells<sup>52</sup>. Interestingly, analysis of gene expression in zebrafish beta cells revealed a subset of cells with high OXPHOS gene expression, of which some had lower insulin gene expression and were interpreted as hub beta cells<sup>53</sup>. Thus, the characteristics of beta cells with high mitochondrial gene expression may be more complex, and differ not only across the life of the individual, but also between species. It may also be of interest to isolate hub cells and assess their correlation with OXPHOS genes expression in the future. Thus, it will be of interest to explore whether the beta cell mitochondrial sub-groups in humans also appear in children, and if they do, whether they correlate with expression of nDNA-encoded mitochondrial genes, as well as with the expression of Insulin.

The identification of the human beta cells subgroups in both healthy and type 2 diabetes individuals support the fundamental importance of such subpopulations of cells for life. As the sample size of humans tested is relatively small, with very few patients and healthy individuals, future increase in sample sizes of such groups is required to draw any conclusion about the implications of our observations to disease conditions.

The positive correlation of mtDNA gene expression in the identified beta cells subgroups with Insulin gene expression lends a first clue for the physiological importance of these subgroups. Specifically, our findings suggest that human beta cells diverge into functionally different groups already at the gene regulatory level, and not only physiologically<sup>17</sup>. Nevertheless, it still remains to be found whether the nature of the subgroups and their composition will change upon exposure to mitochondria-related environmental conditions.

## Online Methods

**Available samples for analysis.** scRNA-seq data from mouse and human pancreatic beta cells were obtained from five studies. Datasets were downloaded from the following sites;

Human and mouse Dataset I: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84133>

Human Dataset II: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81608>

Human Dataset III: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86473>

Human Dataset IV: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85241>

Mouse Dataset II: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77980>

Mouse Dataset III: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86479>

Mouse Dataset IV: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110648>

**Processing of scRNAseq data.** For Dataset I containing human and mouse RNA-seq data, the bioinformatics pipeline of the data processing was carried out as previously reported<sup>22</sup>.

For Datasets II and III of human and mouse sequenced reads were trimmed using Trim Galore (version 0.4.5; [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) while employing default parameters, in addition to the following parameters: [-clip\_R1 n] and [-three\_prime\_clip\_R1 n] (n—representing 5% of the read length) to avoid low quality bases and potential adapter contamination. Trimmed reads were mapped against the reference human genome (GRCh38 for human cells and GRCm38 for mouse cells) using STAR (version 2.5.3)<sup>54</sup>. Mapping of the sequencing reads in the human datasets was performed using default parameters, in addition to the [-outFilterMultimapNmax 1] parameter, to achieve unique mapping, as previously performed<sup>55</sup> to avoid contamination of expressed mitochondrial pseudogenes—mtDNA fragments that were transferred to the nucleus during the course of evolution (NUMTs, see dedicated section below)<sup>56</sup>. Non-unique mapping was also performed for comparison and assessment of such potential contamination. As certain mouse NUMTs are longer and more similar to the active mtDNA, sequencing reads from six mtDNA genes were erroneously filtered out while applying the unique mapping protocol. To overcome such problem, sequencing reads from the mouse datasets were mapped solely against the mtDNA genome using *bwa* with *aln* parameter (BWA-backtrack algorithm)<sup>57</sup>; this enabled subsequent analysis for all mtDNA encoded-genes. Expression levels of all genes were counted using HTSeq-count v0.11.2<sup>58</sup>, using default parameters and employing the [-f bam] parameters. For quality control filtering, gene count values as defined by HTSeq-count were concatenated into a resulting gene expression matrix for each library, which then was loaded into Seurat R-package (version 3.0.2) for subsequent computational analysis. Seurat objects were created using the function “CreateSeuratObject”<sup>25</sup>. Human Dataset IV reads were processed using UMI-tools<sup>59</sup>, which enabled read mapping by STAR (unique mapping), removal of duplicate reads and generation of a gene expression matrix. For all datasets (human and mouse), cell types identities were already reported in the original studies. Nevertheless, we verified such using “FindVariableFeatures” function and clustering in Seurat (shown is a representative analysis in Dataset IV-Fig S11). Notably, quality control filtering of cells and genes was performed, while, using only cells having at least 3000 detected transcripts, with a maximum of 20% ribosomal genes; cells with zero mtDNA read counts were excluded. For all datasets cell doublets were excluded (i.e., cells that were assigned to a given cell type—beta or alpha cell, yet express a mixture of cell type-specific markers—such as both Glucagon and Insulin) (Fig S1, S2, S11). Cells with mtDNA read counts which either exceeded two-fold above the median (for human Dataset I), or displayed more than 10% mtDNA reads were excluded. These measures were taken since overrepresentation of mtDNA genes expression could either associate with stress, or with cell death<sup>60</sup>.

**Cluster identification using Seurat.** To identify clusters of pancreatic beta cells which share patterns of mitochondrial gene expression, Seurat pipeline was utilized<sup>25</sup>. The data matrices were imported and processed with Seurat R package version 3.0.2. To account for the possibility that individual cell complexity leads to cluster



separation and subsequent reduction in the number of total read counts per cell, we used the “vars.to.regress” parameter in scaling function of Seurat. PCA was performed for each separate individual (for both human and mouse experiments) using the mtDNA-protein coding mRNA genes. Although the mtDNA codes for 37 genes, of which 13 encode essential protein-subunits of the OXPHOS system, 2 rRNA genes (12S, 16S) and 22 tRNA genes, the RNA-seq libraries of all datasets enabled analysis of only longer transcripts, while excluding transcripts with short 3' poly-A (i.e. < 10A) in the inDrops platform, which selected for PolyA+ transcripts (Dataset I)<sup>61</sup>. This limited our analysis to the 13 mtDNA-encoded protein coding genes for the Fluidigm C1 platform and to 9 of the 13 mtDNA-encoded OXPHOS subunits (excluding ND5, ND6, ND4L, ATP8 which have a short polyA tail) in the inDrops platform (Table S6). Although the mtDNA is transcribed in strand-specific polycistrons, it is not obvious that mtDNA transcripts will be expressed in the same levels mainly due to post-transcription processing; therefore, multidimensional clustering was performed. Using the first two principle components as input, density clustering was performed per individual to identify cell groups in the data and t-distributed statistical neighbour embedding (tSNE) to visualize the data. A range of values (0.1–1) were examined to assess differences in mitochondrial gene expression. To gain statistical power, the cells of all individuals were clustered; the percent of cells that consistently retained their group identity was calculated and these cells were used for subsequent analyses. Using further Seurat functionality applications, marker genes for each respective cluster were identified and used for subsequent analysis. The specific markers for each cluster identified by Seurat were determined using the “FindAllMarkers” function, using only highly expressed genes (non-zero genes above 0.25 of cells). Finally, to verify the specificity of the identified cell clusters per individual, per dataset, we compared between the identified clusters the expression of 100 random genes that were resampled 1000 times, in each of the tested samples. Heatmaps were generated using ‘DoHeatmap’ function in Seurat.

**Statistical analyses.** Statistical analysis for categorical groups comparisons was performed by unpaired Wilcoxon test with Jack knife 1000X re-sampling test. The latter was performed to control for comparisons of groups with uneven sample sizes. Mutation repertoire and conservation ratio differences were tested using ANOVA. Differential expression of genes was tested using “negbinom” test for Dataset I and Dataset IV which identifies differentially expressed genes between each couple of groups. In brief, we performed a likelihood ratio test of negative binomial generalized linear models. We used the “bimod” test for Datasets II and III which was developed for measurements from the Fluidigm platform<sup>62</sup>. To control for multiple testing, the results of differential genes' expression was FDR corrected.

**Mitochondrial sequence extraction.** Bam files were indexed using default parameters of Samtools v1.3.1 (index command). To create multiple sequence alignment, we generated pileup files using the Samtools mpileup command (default parameters). In addition, we used the -r MT parameter in order to determine read counts per cell, per-base; to facilitate the usage of this parameter for each studied sample we used a custom-made Python script for each sequenced sample. For each given mtDNA position, with sufficient read coverage that passed our quality control filters (see below), the base frequency was calculated by dividing the number of reads which displayed a certain base by the total read coverage per position.

**Variant quality control and filtering.** We counted base changes (i.e. RNA mutations), only in mtDNA positions covered by at least 400 sequencing reads. A mutation was considered trustworthy only if it was covered by at least two sequencing reads from each direction (e.g., forward and reverse), and if the identified mutation was not in the end of the sequencing read. Secondly, high quality variants per sample, per position were determined if the total coverage of the position was > 400 reads with the mutation represented by > 1% coverage in a given nucleotide position. To avoid errors due to low sequence coverage, only cells with at least 1000 covered mtDNA positions within the protein-coding region were included in the variant analyses. Notably, due to low coverage in the non-coding mtDNA region (D-loop), only mutations in the mtDNA coding region were used for subsequent comparison of mutational repertoire between cells. While considering the mtDNA mutational repertoire, permutation analysis (1000 repeats) was performed by resampling 1000 high quality positions per iteration, per cell. Mutations percentage per subgroup was calculated by summing the variants per subgroup and dividing such by 1000 X number of cells per iteration.

**Identification of personalized sub-group mutations.** To identify mutations that are more prevalent in a certain group of cells as compared to the other cells' group per individual, or per tested condition, frequencies of mutations and RNA heterogeneity percentage (mean plus SD) were determined in each of the cell groups (Table S4). Additionally, candidate inherited mutations were identified such that they were shared between alpha and beta cells isolated from the same individual. For each individual the percent of candidate inherited mutations was determined by dividing the number of such by the total number of mutations, per cell subgroup, per individual.

**Assessing the functional potential of mutations in mtDNA transcripts.** To assess whether RNA mutations occurred randomly throughout the mtDNA, or were subjected to selective constraints, the conservation score averages of all the detected mtDNA positions was compared to random distribution. To this end, 100-way phastCons<sup>63,64</sup> score per human and mouse mtDNA position was downloaded from the UCSC website (<http://genome.ucsc.edu/>), and the average score of all RNA heterogenic positions was calculated for each sample. The scores of random distribution were calculated by sample-specific permutation. For each sample, the original number of detected heterogenic positions was resampled ten thousand times, and the average score of all the resampled positions in each iteration was calculated. Next, the expected random value was calculated by

averaging the score of all iterations. The ratio between the observed score average, and the expected random average, was calculated to compare between the distributions of the two cellular sub-populations, per subject. Sample specific *p* values were calculated based on the permutation scores, as the fraction of iterations that had either lower or higher score averages than the observed average (when the observed-expected ratio was lower or higher than 1, respectively).

**Mitochondrial nDNA pseudogenes (NUMTs) likely did not impact expression differences.** It has been known for some time, that the nDNA harbours a repertoire of mtDNA sequence fragments (NUMTs) that were transferred from the mitochondria during the course of evolution. NUMTs potentially pose an obstacle to mtDNA gene expression assessment, as a subset of RNA reads might originate from NUMTs rather than from the active mtDNA. As a first step to control for such a scenario, we performed both unique and non-unique mapping in the human datasets. Similarly, in *mus musculus* there is a large NUMT covering a substantial part of the mtDNA (~4.5 kb), with high sequence similarity to the corresponding mitochondrial reference genome (99.9% identity across this sequence)<sup>45</sup>. As leaving this mtDNA region out will result in data loss for 6 mtDNA genes in mouse, we used *bwa* mapping only for the mtDNA genome sequence. To test the levels of potential NUMTs in the unique mapping data, the percent of NUMT reads (+/- SD) was calculated per cell, per base (Table S8). In addition, whole genome differential expression analysis further filtered out pseudogenes to avoid noise.

Received: 29 August 2020; Accepted: 21 December 2020

Published online: 11 January 2021

## References

- Kahn, B. B. Type 2 diabetes: When insulin secretion fails to compensate for insulin resistance. *Cell* **92**, 593–596 (1998).
- Brun, T. & Maechler, P. Beta-cell mitochondrial carriers and the diabetogenic stress response. *Biochem. Biophys. Acta.* **1863**, 2540–2549 (2016).
- Wollheim, C. B. & Maechler, P. Beta-cell mitochondria and insulin secretion: Messenger role of nucleotides and metabolites. *Diabetes* **51**(Suppl 1), S37–42 (2002).
- Calvo, S. E. & Mootha, V. K. The mitochondrial proteome and human disease. *Annu. Rev. Genomics Hum. Genet.* **11**, 25–44 (2010).
- Gershoni, M. *et al.* Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biol. Evol.* **6**, 2665–2680 (2014).
- Gershoni, M. *et al.* Coevolution predicts direct interactions between mtDNA-encoded and nDNA-encoded subunits of oxidative phosphorylation complex I. *J. Mol. Biol.* **404**, 158–171 (2010).
- Wallace, D. C. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu. Rev. Biochem.* **76**, 781–821 (2007).
- Barshad, G., Marom, S., Cohen, T. & Mishmar, D. Mitochondrial DNA transcription and its regulation: An evolutionary perspective. *Trends Genet.* **34**, 682–692 (2018).
- Aloni, Y. & Attardi, G. Symmetrical in vivo transcription of mitochondrial DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA* **68**, 1757–1761 (1971).
- Barshad, G., Blumberg, A., Cohen, T. & Mishmar, D. Human primitive brain displays negative mitochondrial-nuclear expression correlation of respiratory genes. *Genome Res.* **28**, 952–967 (2018).
- Fernandez-Vizarrá, E., Enriquez, J. A., Perez-Martos, A., Montoya, J. & Fernandez-Silva, P. Tissue-specific differences in mitochondrial activity and biogenesis. *Mitochondrion* **11**, 207–213 (2011).
- Mulder, H. Transcribing beta-cell mitochondria in health and disease. *Mol. Metab.* **6**, 1040–1051 (2017).
- Silva, J. P. *et al.* Impaired insulin secretion and beta-cell loss in tissue-specific knockout mice with mitochondrial diabetes. *Nat. Genet.* **26**, 336–340 (2000).
- Nicholas, L. M. *et al.* Mitochondrial transcription factor B2 is essential for mitochondrial and cellular function in pancreatic beta-cells. *Mol. Metab.* **6**, 651–663 (2017).
- Jitrapakdee, S., Wutthisathapornchai, A., Wallace, J. C. & MacDonald, M. J. Regulation of insulin secretion: Role of mitochondrial signalling. *Diabetologia* **53**, 1019–1032 (2010).
- Avrahami, D., Klochendler, A., Dor, Y. & Glaser, B. Beta cell heterogeneity: An evolving concept. *Diabetologia* **60**, 1363–1369 (2017).
- Johnston, N. R. *et al.* Beta cell hubs dictate pancreatic islet responses to glucose. *Cell Metab.* **24**, 389–401 (2016).
- Avrahami, D. *et al.* beta-Cells are not uniform after all—Novel insights into molecular heterogeneity of insulin-secreting cells. *Diabetes Obes. Metab.* **19**(Suppl 1), 147–152 (2017).
- Wiederkehr, A. & Wollheim, C. B. Mitochondrial signals drive insulin secretion in the pancreatic beta-cell. *Mol. Cell. Endocrinol.* **353**, 128–137 (2012).
- Lawlor, N. *et al.* Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* **27**, 208–222 (2017).
- Xin, Y. *et al.* RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.* **24**, 608–615 (2016).
- Baron, M. *et al.* A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**(346–360), e344 (2016).
- Dorrell, C. *et al.* Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54**, 2832–2844 (2011).
- Rorsman, P., Ramracheya, R., Rorsman, N. J. & Zhang, Q. ATP-regulated potassium channels and voltage-gated calcium channels in pancreatic alpha and beta cells: Similar functions but reciprocal effects on secretion. *Diabetologia* **57**, 1749–1761 (2014).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Wolf, A. R. & Mootha, V. K. Functional genomic analysis of human mitochondrial RNA processing. *Cell Rep.* **7**, 918–931 (2014).
- Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
- Osorio, D. & Cai, J. J. Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA sequencing data quality control. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btaa751> (2020).
- Davidson, H. W., Wenzlau, J. M. & O'Brien, R. M. Zinc transporter 8 (ZnT8) and beta cell function. *Trends Endocrinol. Metab.* **25**, 415–424 (2014).
- Gutierrez-Aguilar, M. & Baines, C. P. Physiological and pathological roles of mitochondrial SLC25 carriers. *Biochem. J.* **454**, 371–386 (2013).

31. Heron, L., Virsolvy, A., Apiou, F., Le Cam, A. & Bataille, D. Isolation, characterization, and chromosomal localization of the human ENSA gene that encodes alpha-endosulfine, a regulator of beta-cell K(ATP) channels. *Diabetes* **48**, 1873–1876 (1999).
32. Pepin, E. *et al.* Short-chain 3-hydroxyacyl-CoA dehydrogenase is a negative regulator of insulin secretion in response to fuel and non-fuel stimuli in INS832/13 beta-cells. *J. Diabetes* **2**, 157–167 (2010).
33. Stutzer, I., Esterhazy, D. & Stoffel, M. The pancreatic beta cell surface proteome. *Diabetologia* **55**, 1877–1889 (2012).
34. Lipson, K. L. *et al.* Regulation of insulin biosynthesis in pancreatic beta cells by an endoplasmic reticulum-resident protein kinase IRE1. *Cell Metab.* **4**, 245–254 (2006).
35. Xin, Y. *et al.* Pseudotime ordering of single human beta-cells reveals states of insulin production and unfolded protein response. *Diabetes* **67**, 1783–1794 (2018).
36. She, H. *et al.* Direct regulation of complex I by mitochondrial MEF2D is disrupted in a mouse model of Parkinson disease and in human patients. *J. Clin. Invest.* **121**, 930–940 (2011).
37. Muraro, M. J. *et al.* A Single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**(385–394), e383 (2016).
38. Dorrell, C. *et al.* Human islets contain four distinct subtypes of beta cells. *Nat. Commun.* **7**, 11756 (2016).
39. Bar-Yaacov, D. *et al.* RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. *Genome Res.* **23**, 1789–1796 (2013).
40. Safra, M. *et al.* The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251–255 (2017).
41. Bar-Yaacov, D. *et al.* Mitochondrial 16S rRNA is methylated by tRNA methyltransferase TRMT61B in all vertebrates. *PLoS Biol.* **14**, e1002557 (2016).
42. Xin, Y. *et al.* Use of the fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc. Natl. Acad. Sci. USA* **113**, 3293–3298 (2016).
43. Zeng, C. *et al.* Pseudotemporal ordering of single cells reveals metabolic control of postnatal beta cell proliferation. *Cell Metab.* **25**(1160–1175), e1111 (2017).
44. Ruzzenente, B. *et al.* LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J.* **31**, 443–456 (2012).
45. Calabrese, F. M., Simone, D. & Attimonelli, M. Primates and mouse NumtS in the UCSC genome browser. *BMC Bioinform.* **13**(Suppl 4), S15 (2012).
46. Lu, T. T. *et al.* The polycomb-dependent epigenome controls beta cell dysfunction, dedifferentiation, and diabetes. *Cell Metab.* **27**(1294–1308), e1297 (2018).
47. Farack, L. *et al.* Transcriptional heterogeneity of beta cells in the intact pancreas. *Dev. Cell* **48**(115–125), e114 (2019).
48. Schuit, F. *et al.* Metabolic fate of glucose in purified islet cells. Glucose-regulated anaplerosis in beta cells. *J. Biol. Chem.* **272**, 18572–18579 (1997).
49. Quesada, I., Todorova, M. G. & Soria, B. Different metabolic responses in alpha-, beta-, and delta-cells of the islet of Langerhans monitored by redox confocal microscopy. *Biophys. J.* **90**, 2641–2650 (2006).
50. Yoshihara, E. *et al.* ERGamma is required for the metabolic maturation of therapeutically functional glucose-responsive beta cells. *Cell Metab.* **23**, 622–634 (2016).
51. Cabrera, O. *et al.* The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc. Natl. Acad. Sci. USA* **103**, 2334–2339 (2006).
52. Cnop, M. *et al.* The long lifespan and low turnover of human islet beta cells estimated by mathematical modelling of lipofuscin accumulation. *Diabetologia* **53**, 321–330 (2010).
53. Salem, V. *et al.* Leader beta-cells coordinate Ca(2+) dynamics across pancreatic islets in vivo. *Nat. Metab.* **1**, 615–629 (2019).
54. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
55. Cohen, T., Levin, L. & Mishmar, D. Ancient out-of-africa mitochondrial DNA variants associate with distinct mitochondrial gene expression patterns. *PLoS Genet.* **12**, e1006407 (2016).
56. Mishmar, D., Ruiz-Pesini, E., Brandon, M. & Wallace, D. C. Mitochondrial DNA-like sequences in the nucleus (NUMTs): Insights into our African origins and the mechanism of foreign DNA integration. *Hum. Mutat.* **23**, 125–133 (2004).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
58. Anders, S., Pyl, P. T. & Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
59. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
60. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
61. Slomovic, S., Laufer, D., Geiger, D. & Schuster, G. Polyadenylation and degradation of human mitochondrial RNA: The prokaryotic past leaves its mark. *Mol. Cell. Biol.* **25**, 6427–6435 (2005).
62. McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013).
63. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
64. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

## Acknowledgements

This study was funded by the Israel Science Foundation Grant Number 372/17 and by the US Army life sciences division Grant Number LS67993 awarded to DM. The authors thank Dr. Tal Shay (BGU) for critical discussion during early stages of manuscript preparation.

## Author contributions

H.M. wrote scripts, analyzed the data, drew the figures and participated in writing the paper; T.C. participated in data analysis; D.M. conceived the study, supervised the analyses and wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80334-w>.

**Correspondence** and requests for materials should be addressed to D.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2021