**Statistical Methods**

Sequencing was conducted on a NextSeq 1000 Sequencer (Illumina). Generation of FASTQ files, trimming of reads, alignment to a reference genome (generation of .bam files) and quantification of reads was automatically performed within the DRAGEN analysis software (v. xxx, Illumina). The generated count matrix served as input for the analysis in R (v. 4.3.2 or higher) following the DESeq2 pipeline (<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>).

Data quality and potential presence of batch effects were assessed via exploratory data analysis including principal component analysis (PCA), scatterplot matrices (R package *GGally,* https://ggobi.github.io/ggally/authors.html) and dispersion plots. Genes with a dispersion parameter >1 were investigated more closely to assess whether dispersion is attributable to technical or biological variability. Cooks distance was used to check for outliers.

Differential expression was calculated with the DESeq() function comparing melanocytes with dermal stem cells (DSCs) at a significance level of α = 0.05 (FDR-adjusted) for a 1.5 fold change (0.585 log fold change) difference. The negative binomial model included only the cell type. Raw read counts were provided as the input, with each column representing a sample and each row representing a gene. For all analyses DSCs were the *control* group whereas Melanocytes were the *treated* group. Log fold changes were shrunk using the lfcshrink() function with the apeglm algorithm provided by <https://pubmed.ncbi.nlm.nih.gov/30395178/> due to high variability in genes with low counts. Counts were transformed with the regularized log approach (rlog) as described in Love et. al 2014 with the argument blind = FALSE. Differentially expressed genes were examined via Volcano plot (BioConductor package *EnhancedVolcano,* https://github.com/kevinblighe/EnhancedVolcano), barplots of selected marker genes and hierarchical clustering (BioConductor package *ComplexHeatmap,* https://pubmed.ncbi.nlm.nih.gov/27207943/). All annotations were based on the homo sapiens annotation retrieved from the BioConductor package org.Hs.eg.db (v. 3.19.1, <https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html>)

To investigate the biological function of differentially expressed genes, functional analyses with the R package pathfindR (<https://cran.r-project.org/web/packages/pathfindR/vignettes/intro_vignette.html>) were conducted. This package includes protein-protein interaction information harvested from active subnetworks for a active-subnetwork-oriented pathway analysis, which yields superior results compared to conventional enrichment analyses (<https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00858/full>). The gene set database used in this work was WikiPathways. The run\_pathfindR() function only includes GO or KEGG analysis by default. However, it allows for supply of a custom gene set. Therefore, we loaded the WikiPathways gene set as .gmt file with the BioConductor package qusage (<https://www.bioconductor.org/packages/release/bioc/html/qusage.html>) after downloading the most recent release (wikipathways-20240510-gmt-Homo\_sapiens.gmt) from the WikiPathways homepage (<https://www.wikipathways.org/download.html>) and supplied it as custom gene set. A reduced background gene list (as suggested for RNASeq in <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009935>) for the analysis contained all genes which were expressed above a minimum read threshold of 10 counts in at least 5 samples or which were differentially expressed. Shrunken log2 fold changes and fdr-adjusted p-values were provided as input.

Following the enrichment analysis, overrepresented pathways were scored with the pathfindR function score\_terms() to assess whether pathways were activated or repressed. Activated and repressed pathways were separated into groups and enriched functional categories were clustered in each group, respectively. These clusters were plotted as dot plots. Furthermore, representative pathways (the most significantly enriched pathway from each cluster) were extracted and ranked by taking the square root of the product of fold-enrichment and negative log10 of the lowest enrichment p-value. The depiction of enriched representative pathways was conducted in a word cloud where pathway size was directly correlated to the assigned rank.

Additionally, genes with a high dispersion (dispersion >1) were subject to a pathway analysis to identify processes which differ between donors. As input, we only had a list of genes without fold changes or p-values, which made the previous approach with pathfindR unfeasible. Therefore, enrichment analysis has been conducted with the BioConductor package goseq (<https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-2-r14>), a form of overrepresentation analysis optimized for rnaseq by taking transcript lengths into account to adjust for transcript length bias. The input were all genes with a minimum count of 10 in at least 5 samples, encoded with a 1 (high dispersion gene) or 0 (other genes). Enrichment was conducted for GO biological processes and redundancy of enriched terms was reduced by only keeping the parent term and removing related terms. Similarity was calculated with the rrvgo (<https://bioconductor.org/packages/release/bioc/html/rrvgo.html>) function calculateSimMatrix() and based on the output, terms were reduced with the reduceSimMatrix() function and a similarity threshold of 0.7. The reference genome for the goseq() function was “hg19”. Results were depicted in a word cloud showing the top30 enriched GO terms based on the negative log10 p-value of the goseq output after fdr-adjustment.

Master regulators for the comparison of Melanocytes vs DSCs (*controls*) have been identified with the *corto* R package (<https://pubmed.ncbi.nlm.nih.gov/32232425/>) using 294 centroids (transcription factors and regulators) specified in the package as well as using 330 centroids (additional inclusion of known Melanocyte, DSC and Fibroblast markers and transcription factors). The p-value threshold for edge generation was p = 0.0001 and the Top10 up- or down-regulated master regulators were plotted with the mraplot() function.

Functional analyses were performed with the BioConductor package *clusterProfiler* (https://pubmed.ncbi.nlm.nih.gov/34557778/) to identify enriched KEGG pathways and gene ontology (GO) terms. Redundant GO terms were removed from the analysis based on semantic similarity utilizing the simplify() function with the arguments cutoff = 0.7, by = “p.adjust” and select\_fun = min of the *clusterProfiler* package. Enriched pathways were depicted as dotplots containing the Top20 enriched pathways and complex gene-pathway interactions were illustrated by the cnetplot() function to draw networks of genes (with color-coded log fold changes) and associated pathways. Either the Top5 enriched pathways were drawn as a network (to retain interpretability) or enriched pathways associated with melanocyte development or pigmentation were shown.