



## Article

# Cell-Free Genic Extrachromosomal Circular DNA Profiles of DNase Knockouts Associated with Systemic Lupus Erythematosus and Relation with Common Fragile Sites

Daniela Gerovska <sup>1</sup> , Patricia Fernández Moreno <sup>1</sup>, Aitor Zabala <sup>1</sup> and Marcos J. Araúzo-Bravo <sup>1,2,3,\*</sup>

- <sup>1</sup> Computational Biology and Systems Biomedicine, Biogipuzkoa Health Research Institute, Calle Doctor Begiristain s/n, 20014 San Sebastian, Spain; daniela.gerovska@biodonostia.org (D.G.); patricia.fernandezmoreno@biodonostia.org (P.F.M.); aitor.zabalagarcia@biodonostia.org (A.Z.)
- <sup>2</sup> Basque Foundation for Science, IKERBASQUE, Calle María Díaz Harokoa 3, 48013 Bilbao, Spain
- <sup>3</sup> Department of Cell Biology and Histology, Faculty of Medicine and Nursing, University of Basque Country (UPV/EHU), 48940 Leioa, Spain
- \* Correspondence: mararabra@yahoo.co.uk or marcos.arauzo@biodonostia.org; Tel.: +34-943-006108

**Abstract:** Cell-free extrachromosomal circular DNA (cf-eccDNA) has been proposed as a promising early biomarker for disease diagnosis, progression and drug response. Its established biomarker features are changes in the number and length distribution of cf-eccDNA. Another novel promising biomarker is a set of eccDNA excised from a panel of genes specific to a condition compared to a control. Deficiencies in two endonucleases that specifically target DNA, Dnase1 and Dnase113, are associated with systemic lupus erythematosus (SLE). To study the genic eccDNA profiles in the case of their deficiencies, we mapped sequenced eccDNA data from plasma, liver and buffy coat from *Dnase1* and *Dnase113* knockouts (KOs), and wild type controls in mouse. Next, we performed an eccDNA differential analysis between KO and control groups using our DifCir algorithm. We found a specific genic cf-eccDNA fingerprint of the Dnase113 group compared to the wild type controls involving 131 genes; 26% of them were associated with human chromosomal fragile sites (CFSs) and with a statistically significant enrichment of CFS-associated genes. We found six genes in common with the genic cf-eccDNA profile of SLE patients with DNASE1L3 deficiency, namely *Rorb*, *Mvb12b*, *Osbpl10*, *Fto*, *Tnik* and *Arhgap10*; all of them were specific and present in all human plasma samples, and none of them were associated with CFSs. A not so distinctive genic cf-eccDNA difference involving only seven genes was observed in the case of the Dnase1 group compared to the wild type. In tissue—liver and buffy coat—we did not detect the same genic eccDNA difference observed in the plasma samples. These results point to a specific role of a set of genic eccDNA in plasma from *DNase* KOs, as well as a relation with CFS genes, confirming the promise of the genic cf-eccDNA in studying diseases and the need for further research on the relationship between eccDNA and CFSs.



**Citation:** Gerovska, D.; Fernández Moreno, P.; Zabala, A.; Araúzo-Bravo, M.J. Cell-Free Genic Extrachromosomal Circular DNA Profiles of DNase Knockouts Associated with Systemic Lupus Erythematosus and Relation with Common Fragile Sites.

*Biomedicines* **2024**, *12*, 80. <https://doi.org/10.3390/biomedicines12010080>

Academic Editor: Mikhail Kostik

Received: 28 November 2023

Revised: 25 December 2023

Accepted: 26 December 2023

Published: 28 December 2023

**Keywords:** extrachromosomal circular DNA (eccDNA); systemic lupus erythematosus (SLE); Dnase1; Dnase113; chromosomal fragile site (CFS); genome-wide association study (GWAS); DifCir; produced per gene circle (PpGC); differentially produced per gene circle (DPpGC); plasma; liver; buffy coat



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

DNA degradation is critical for healthy organism development and survival. The regular development and healthy function of humans and mice requires two families of endonucleases, Dnase1 and Dnase2. Deoxyribonuclease 1, Dnase1, and deoxyribonuclease 1 like 3, Dnase113, belong to the Dnase1 family, with Dnase1 expressed primarily in the digestive tract and in blood, where it helps clean free DNA from the circulation. Dnase113 is expressed by dendritic cells and macrophages in the liver and spleen, in myeloid cells in blood, and in neurons, with an unknown role in the latter [1]. A disease associated with deficiencies of Dnase1 and Dnase113 in mouse is systemic lupus erythematosus

(SLE). SLE is characterized by antibody generation, activation and complement by these antibodies, inflammation, and organ damage, with the secreted DNases, Dnase1 and Dnase113, providing the greatest protection from autoantibody formation [1].

Cell-free (cf) extrachromosomal circular DNA (eccDNA) has a potential clinical application as a biomarker in cancer [2] and other diseases such as diabetes [3], inflammatory bowel disease [4], and amyotrophic lateral sclerosis [5], among others. EccDNA features such as changes in the number of the eccDNAs, changes in the length frequency distributions of the eccDNA or differential abundance of eccDNA from certain genes [3,6,7] have been considered as biomarkers. Previous work studying the relationship of nucleases and eccDNA characteristics in plasma using knockout (KO) mouse models with a deficiency in Dnase1 or Dnase113 found that cf-eccDNA *Dnase113*<sup>-/-</sup> exhibited larger size distributions than in wild type mice [8]. Anyway, current studies still suffer from a limited number of patients and samples to validate the statistical power of such features to discriminate between conditions.

Common fragile sites (CFSs) are specific genomic loci that are difficult to replicate and correspond to regions of genomic instability. Upon replication stress, CFSs form gaps or breakages, a phenomenon called CFS expression, which correlate positively with chromosomal rearrangement and copy number variation, and where under-replicated DNA subsequently undergoes mitotic DNA synthesis, MiDAS [9]. Based on their frequency, CFSs encompass two classes: the rare CFSs, which are present in the 5% of the population, and common CFSs present in nearly all the population. Intrinsic characteristics of the CFSs are the overlap with very large and active genes, with many of these large genes involved in neurological development [10,11]. The CFSs are known to be conserved features of human and mouse chromosomes [11]. CFSs have been extensively studied due to their important pathophysiological relevance in diseases such as cancer and neurological disorders, for reviews of CFSs see [12,13]. The possible relationship between CFSs and eccDNA, and CFSs and SLE has not been studied so far.

Here we explored the possibility that the DNase KO groups differ from the wild type (WT) group in their genic eccDNA profiles in mouse plasma and tissue samples, as in the case of cf-eccDNA from SLE patients with *DNASE1L3* mutations [14]. We performed a circular differential analysis based on the split read signal of the eccDNA using our DifCir algorithm [15] of circulomics data from WT, *Dnase113*<sup>-/-</sup> (from here on denoted as Pla-KO113) and *Dnase1*<sup>-/-</sup> (here denoted as Pla-KO1) eccDNA purified plasma samples, comparing the KOs with the WT (here denoted as Pla-WT), control samples. We compared the obtained top-ranked genes producing differential eccDNA in plasma with the eccDNA produced by the same genes in KO and healthy control tissue—buffy coat and liver—samples. We aimed to look for commonalities between the genic eccDNA profiles of the two SLE mouse models and humans monogenic for SLE with *DNASE1L3* deficiency. Furthermore, we explored the relationship between genic eccDNA production and the excision of the eccDNA from CFSs in both human SLE and mouse models.

## 2. Methods

### 2.1. Samples and Data

Plasma DNA samples from 12 WT, 11 *Dnase1*<sup>-/-</sup>, and 11 *Dnase113*<sup>-/-</sup> mice [8] were used to study the genic cell-free (cf) eccDNA profiles of the two DNase KOs compared to the WT. For the plasma samples, we used short-read sequenced data of eccDNA libraries obtained with a tagmentation (Tn)-based preparation protocol [16].

Additionally, we used 5 WT, 5 *Dnase1*<sup>-/-</sup> and 5 *Dnase113*<sup>-/-</sup> liver, and 6 WT, 4 *Dnase1*<sup>-/-</sup>, and 5 *Dnase113*<sup>-/-</sup> buffy coat samples to assess the eccDNA excising behavior of the genes by characterizing the genic cf-eccDNA profiles of the KOs in liver and buffy coat tissues. For the tissue samples, we used short-read sequenced data of eccDNA libraries obtained using both a Tn- and a rolling circle amplification (RCA)-based preparation protocol [16].

All data have been downloaded from the European Genome-Phenome Archive (EGA) hosted by the European Bioinformatics Institute (EBI) (accession EGAS00001005873) [8].

## 2.2. Mapping of eccDNA and Quantification of Produced per Gene Circles/eccDNAs (PpGCs)

Mapping and quantification of produced per gene eccDNAs (PpGCs) was performed as described previously [15]. EccDNA reads were mapped with Circle\_Finder [17,18] using as parameters the mm10 build of the mouse genome and a minNonOverlap threshold between two split reads of 10 bp. Clusters of eccDNAs within a distance smaller than  $D_{min} = 10$  bp were coalesced and the number of the split reads detecting them were summed as split reads of the merged eccDNAs. EccDNAs with less than two split reads were excluded. EccDNA annotations were accomplished with bedtools intersect using Genecode gene coordinates. EccDNA sizes were assessed from mitochondria-free DNAs, setting 10 Kbp as the cutoff for all small size circles. Next, the number of split reads of all those eccDNAs that carried the same gene or fragment of a gene were added to obtain the unscaled produced per gene eccDNA (PpGC<sub>*i*</sub>) sum for each gene *i*. For scaling by gene length, each PpGC<sub>*i*</sub> was multiplied by a scale factor  $L_{Max}/L_i$ , where  $L_{Max}$  is the length of the longest gene found in the dataset, and  $L_i$  is the length of the gene *i*. Finally, data equalization was performed using the  $\log_2(\text{PpGC} + 1)$  transformation of the scaled PpGCs to obtain the final PpGCs.

## 2.3. Identification of Differentially Produced per Gene eccDNAs (DPpGCs)

We used our method—DifCir—for the differential analysis of sequenced purified eccDNA data based on their split read signal [15]. We calculated the mean values of the PpGCs for each group of replicates and selected as differentially produced per gene eccDNAs (DPpGCs) those whose absolute difference of the means between the two groups was more than a selection threshold  $\theta_{\text{DPpGC}} = \text{one-fold change (FC)}$  on the  $\log_2$  scale. Finally, we selected the statistically significant DPpGCs based on Student's t-test using a significance threshold of  $\alpha_{\text{DPpGC}} = 0.05$ .

## 2.4. Calculation of the Chromosome Enrichment of DPpGCs

The chromosome enrichment of the DPpGCs is calculated as the percentage of the difference of the ratio of DPpGCs in each chromosome minus the ratio of DPpGCs in the whole genome as follows:  $\Delta c = 100(Nd_c/N_c - Nd_g/N_g)$ , where  $Nd_c$  is the number of DPpGCs of chromosome *c*,  $N_c$  is the number of eccDNAs of the chromosome *c*,  $Nd_g$  is the number DPpGCs in the whole genome, and  $N_g$  is the number of eccDNAs in the whole genome. The  $\Delta c$  value shows the percentage of enrichment or depletion (for negative cases) of DPpGCs of each chromosome with respect to the percentage of DPpGCs in the whole genome. Since a high percentage of difference  $\Delta c$  does not ensure statistical significance, to assess the significance of the results, the hypergeometric test was used with mid-*p*-values. The significantly enriched chromosomes are labeled with stars \*, where \* means that  $0.01 \geq \text{mid-} p\text{-value} > 0.001$ , \*\* means that  $0.001 \geq \text{mid-} p\text{-value} > 0.0001$ , etc. A higher number of asterisks corresponds to higher statistical significance of the results.

## 2.5. Finding Common Genes between up-DPpGCs in Plasma from *Dnase1l3*<sup>-/-</sup> Mice and SLE Patients with DNASE1L3 Deficiency

We intersected the list of genes identified as statistically significant up-DPpGCs in the mouse *Dnase1l3*<sup>-/-</sup> group compared to WT here with the list of 267 statistically significant (significance level  $\alpha_{\text{DPpGC}} = 0.01$ ) in the SLE patients with DNASE1L3 deficiency (Pla-SLE) compared to a healthy control group (Pla-HC) from [14].

## 2.6. Finding Common DPpGCs with Chromosomal Fragile Site (CFS) Genes

The CFS genes were taken from the database of fragile sites in human chromosomes (HumCFS) [19], downloaded on 6 November 2023. The CFSs were stratified in 6 categories based on the drug used to destabilize the DNA (5-azacytidine, aphidicolin, BrdU, distamycinA, folic acid, and the pool of the five drugs). For each of the CFSs, we collected the genes located in each CFS locus and annotated them with the corresponding drug associated with the CFS. To calculate the statistical enrichment of the CFS genes, we used

the sixth group constituted by pooling all of the CFS genes affected by any of the CFS drugs. To assign the original human genes names to mouse genes names, we used the NCBI HomoloGene database. The statistical significance of the enrichment of the CFSs was interrogated using the hypergeometric test using a significance threshold of  $\alpha_{\text{CFS}} = 0.05$ .

### 2.7. Finding Common DPpGCs with GWAS for 'Systemic Lupus Erythematosus'

The Genome-Wide Association Study (GWAS) data were downloaded on 5 February 2023 from the GWAS Catalog of the National Human Genome Research Institute (NHGRI) and European Bioinformatics Institute (EBI). The data in the GWAS Catalog used are currently mapped to human genome assembly GRCh38.p13 and dbSNP Build 154. The statistical significance of the enrichment of any of the diseases in the GWAS Catalog was interrogated using the hypergeometric test with a significance threshold of  $\alpha_{\text{GWAS}} = 0.05$ .

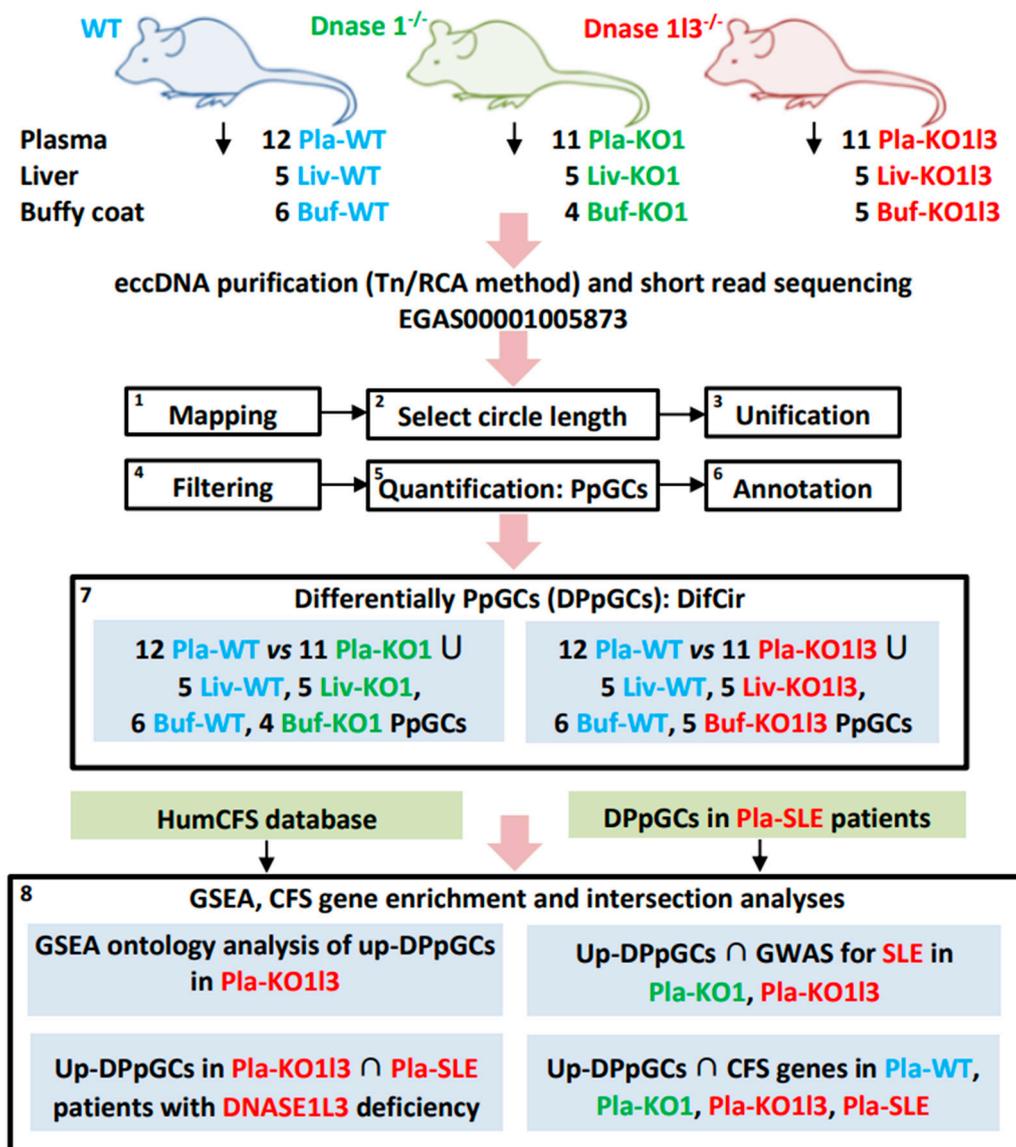
## 3. Results

### 3.1. Workflow

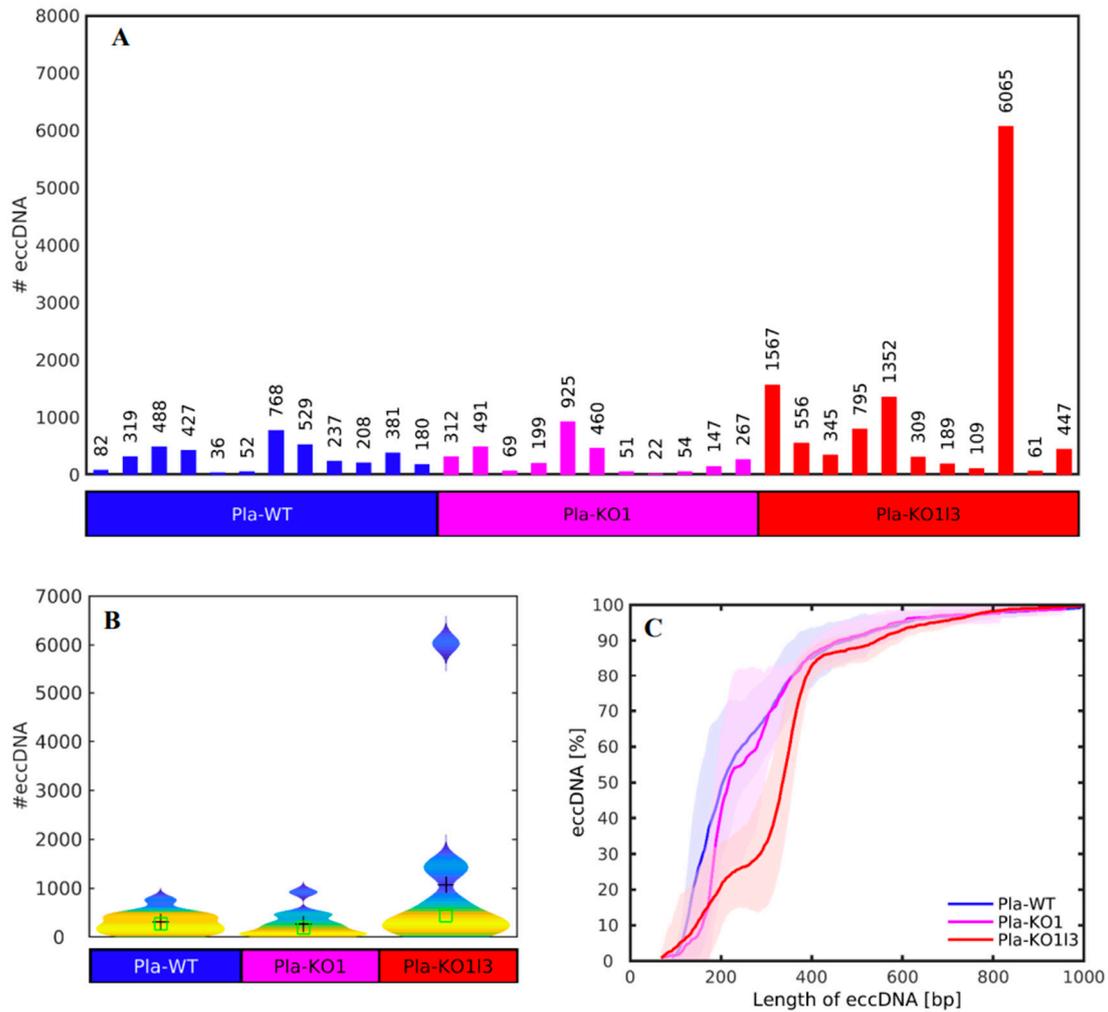
We aimed to obtain and analyze the genic cf-eccDNA profiles of *Dnase1* and *Dnase113* KOs compared to wild type mice (WT). While Sin et al. [8] asked the research question what was the effect of the nucleases on the eccDNA properties like their abundance and size distribution, the question about possible differences in the way they affect the excision of eccDNA from different genes remained. Here, we processed circulomics data of cell-free (cf)-eccDNA from Pla-WT,  $n = 12$ , Pla-KO1 (*Dnase1*<sup>-/-</sup>),  $n = 11$ , and Pla-KO113 (*Dnase113*<sup>-/-</sup>),  $n = 11$ , groups from Sin et al. [8]. EccDNA has been isolated, purified and short-read sequenced from plasma, liver and buffy coat of WT mice, mice with *Dnase1* knockout (KO1), and mice with *Dnase113* knockout (KO113). EccDNA libraries from plasma samples were constructed using a tagmentation (Tn)-based method, while for the tissue samples of liver and buffy coat, both rolling circle amplification (RCA) and Tn have been used [8]. This was followed by mapping, annotation, and quantification of eccDNAs, and quantification of produced per gene circles (PpGCs) for the plasma, liver and buffy coat samples. Next, the calculation of differentially PpGCs (DPpGCs) was performed for Pla-KO1 compared to Pla-WT and for Pla-KO113 compared to Pla-WT. The statistically significant up-DPpGCs in the two directions in the two comparisons were cross-checked with the PpGCs in liver and buffy coat obtained through the Tn and RCA methods. Functional enrichment analysis of the up-DPpGCs in Pla-KO113 vs. Pla-WT was performed with the gene sets from GSEA. To study the extent to which the mouse model reflects the human SLE disease, a comparison of the up-DPpGCs in Pla-KO113 vs. Pla-WT and the up-DPpGCs in SLE patients with DNASE1L3 deficiency vs. healthy controls has been implemented. Additionally, the intersection with the term 'systemic lupus erythematosus' from the GWAS database was performed to annotate in our results the genes that are already associated with SLE. The up-DPpGCs in the four comparisons were checked for genes from chromosomal fragile sites (CFSs) and annotated as such, and additionally according to the agent used to induce the CFSs. The data and schematic workflow of this work are presented in detail in Figure 1.

### 3.2. The Number of cf-eccDNAs Does Not Discriminate between the DNase KOs and WT

First, we studied whether the number or the sizes of the cf-eccDNAs could be used to discriminate between the two DNase KOs and the WT. The number and the length distribution of the cf-eccDNAs mapped by us are presented in Figure 2. We observed no significant differences between the number (Figure 2A) and size distribution (Figure 2B) of eccDNAs in Pla-WT and Pla-KO1 samples. The Wilcoxon rank sum test confirmed that the 1.1338-fold difference is not statistically significant between Pla-KO1 (mean  $272.5 \pm \text{sem } 81.6$ ), where sem is the standard error of the mean, and Pla-WT (mean  $308.9 \pm \text{sem } 63.6$ ) at 1% significance level with a  $p$ -value 0.5254.



**Figure 1.** Experimental setup and computational analysis workflow of the cell-free and tissue circulomics data from *Dnase1*<sup>-/-</sup>, *Dnase113*<sup>-/-</sup>, and wild type (WT) mice from Sin et al. [8]. The computational analysis includes mapping, annotation, and quantification—using PpGCs—of eccDNAs, in plasma, liver, and buffy coat, followed by the calculation of DPpGCs for Pla-KO1 compared to Pla-WT and for Pla-KO113 compared to Pla-WT. The statistically significant up-DPpGCs in the two directions in the two comparisons were cross-checked with the PpGCs in liver and buffy coat obtained through the Tn and RCA methods, and annotated for genes from CFSs and according to the agent used to induce the CFSs, followed by calculation for CFS gene enrichment. Additionally, a functional enrichment analysis of the up-DPpGCs in Pla-KO113 vs. Pla-WT was performed with GSEA. A comparison between the up-DPpGCs in Pla-KO113 and in SLE patients with DNASE1L3 deficiency was implemented. Blue, green and red colors mark WT, Pla-KO1 and Pla-KO111 (and patients with Pla-SLE with DNASE1L3 deficiency), respectively.



**Figure 2.** Number (#) and length distribution of the unique eccDNAs mapped in the three mouse plasma sample groups. **(A)** Number of unique eccDNAs. **(B)** Violin plots of the number of eccDNAs. The black crosses and the green squares mark the position of the mean and the median of the distributions, respectively. **(C)** Cumulative frequency distribution of the lengths of the eccDNAs in the Pla-WT (wild type). The shadowed blue, magenta, and red regions cover the standard deviation of the enrichment across the Pla-WT, Pla-KO1 and Pla-KO13 groups, respectively. Pla-WT,  $n = 12$ , Pla-KO1 ( $Dnase1^{-/-}$ ),  $n = 11$ , and Pla-KO13 ( $Dnase113^{-/-}$ ),  $n = 11$ , groups.

In the case of Pla-KO13, there is a higher variability in the number of eccDNAs than in the other conditions, with a sample with an extremely high number of 6065 eccDNAs (Figure 2A); however, the difference in relation to Pla-WT is not statistically significant, and the sample with a high number of eccDNAs could be considered as an outlier based on the violin plots of the distributions of the numbers of eccDNAs (Figure 2B). The Wilcoxon rank sum test did not indicate that the 0.2881-fold difference between Pla-KO13 ( $1072.3 \pm 520.7$ ) and Pla-WT is statistically significantly different at a 1% significance level with a  $p$ -value 0.1507. Neither indicate that the 0.2541-fold difference in the number of eccDNAs in Pla-KO13 and Pla-KO1 is statistically significant at a 1% significance level with a  $p$ -value 0.0759.

In the case of the other classical eccDNA marker, the length of the eccDNAs, the percentages of eccDNAs with lengths ranging from 0 to 1000 bases is very similar for Pla-WT and Pla-KO1 (Figure 2C). In the case of Pla-KO13, there is a slight decrease in plasma KO13 eccDNAs with lengths ranging from 200 to 400 bases. However, the percentages of lengths become similar for the three analyzed cases for eccDNAs with lengths greater than 400 bases. Altogether, Figure 2, as well as a previous detailed analysis of these features [8], show that the length and frequency do not discriminate confidently between the groups.

### 3.3. Existing but Scarce Genic cf-eccDNA Differences between *Dnase1* KO and WT

The traditional approaches, numbers and lengths of eccDNAs, to discriminate between sample groups using circulomics data did not find differences between *Dnase1* KO and WT in plasma samples. Such techniques are not very sensitive since they pool all the potential information of the eccDNAs in a global number for each sample; hence, we developed our DifCir method based on gene-centric information, produced per gene circles (PpGCs). Therefore, to assess whether some genes in the mouse genome were particularly prone to form eccDNA in the *Dnase1* KO group compared to the WT group, we calculated the PpGCs for the Pla-KO1 and Pla-WT samples. The paired scatter plot in Figure 3A shows certain differences in PpGCs in the Pla-KO1 vs. Pla-WT groups, and the volcano plot in Figure 3B shows that some of these differences are statistically significant. For a  $\log_2$  fold change  $\geq 1$  and significance level threshold  $\alpha_{\text{DPpGC}} = 0.05$ , we identified 13 up-DPpGCs in Pla-WT compared to Pla-KO1 (*Gm30211*, *Igkj1-4*, *Slc14a2*, *Adcy2*, *Kcnq5*, *Gm49171*, *Hivep3*, *Myo3b*, *Kcnh8* and *Slc22a22*), as depicted in the heatmap of Figure 3C, and 7 up-DPpGCs in Pla-KO1 related to Pla-WT (*Tshz2*, *Bicra*, *Tnni3k*, *Cfap61*, *Palld*, *Kirrel3*, and *Cdh4*), as depicted in the heatmap of Figure 3D. Among the latter, *Tshz2*, teashirt zinc finger homeobox 2, is a tumor suppressor that inhibits metastasis [20], and *Bicra*, BRD4 interacting chromatin remodeling complex associated protein, enables transcription regulator activator activity. To assess the relation between the DPpGCs and chromosomal fragile sites (CFSs), we marked in the heatmaps the gene names associated with CFSs confirmed by different drugs. We found two CFS-associated genes *Myo3b* (aphidicolin-induced CFS) and *Kcnq5* (BrdU-induced CFS) in Pla-WT and another two, *Tnni3k* (aphidicolin-induced CFS) and *Cfap61* (folic acid-induced CFS) in Pla-KO1, i.e., 15% and 29% of the up-DPpGCs, respectively. However, for the comparisons in the two directions, the number of DPpGCs in CFSs is not statistically significantly enriched (Table 1). Furthermore, to track and compare the production of eccDNA in tissue with available circulomics data for the up-DPpGCs in Pla-WT and Pla-KO1, we represented in the heatmaps the PpGCs in liver and buffy coat tissue from the WT and KO1 samples for the libraries obtained with the RCA method (Figure 3C,D). We observed that the up-DPpGCs in the Pla-WT vs. Pla-KO1 and Pla-KO1 vs. Pla-WT groups do not correlate with the PpGCs in liver and buffy coat, where the PpGCs for the genes up-producing cf-eccDNA in the two comparisons have generally high values in both WT and KO liver and buffy coat samples. We chose to show in the heatmaps in Figure 3C,D the PpGCs in tissue samples from libraries obtained with RCA method since using the RCA method for tissues led to the mapping of a high number of eccDNAs, while the Tn method resulted in mapping of very few eccDNAs. The same heatmaps with PpGCs from liver and buffy coat obtained with the Tn method are represented in Supplementary Figure S1A,B.

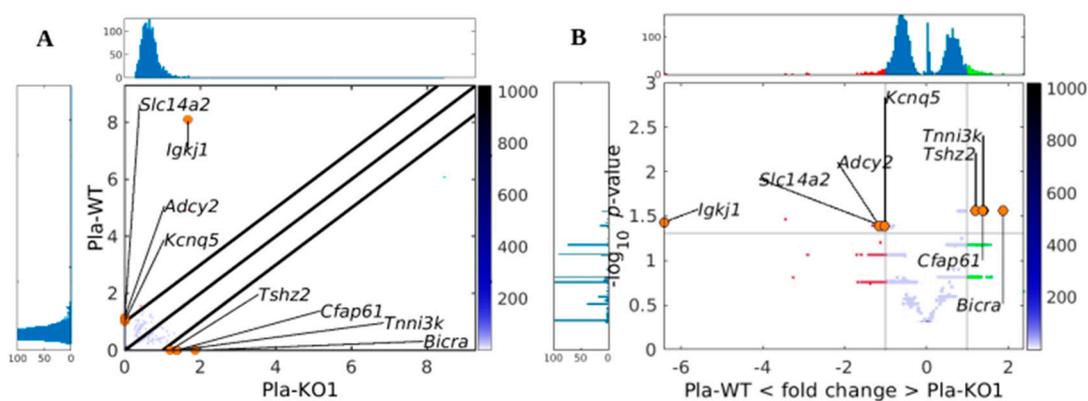
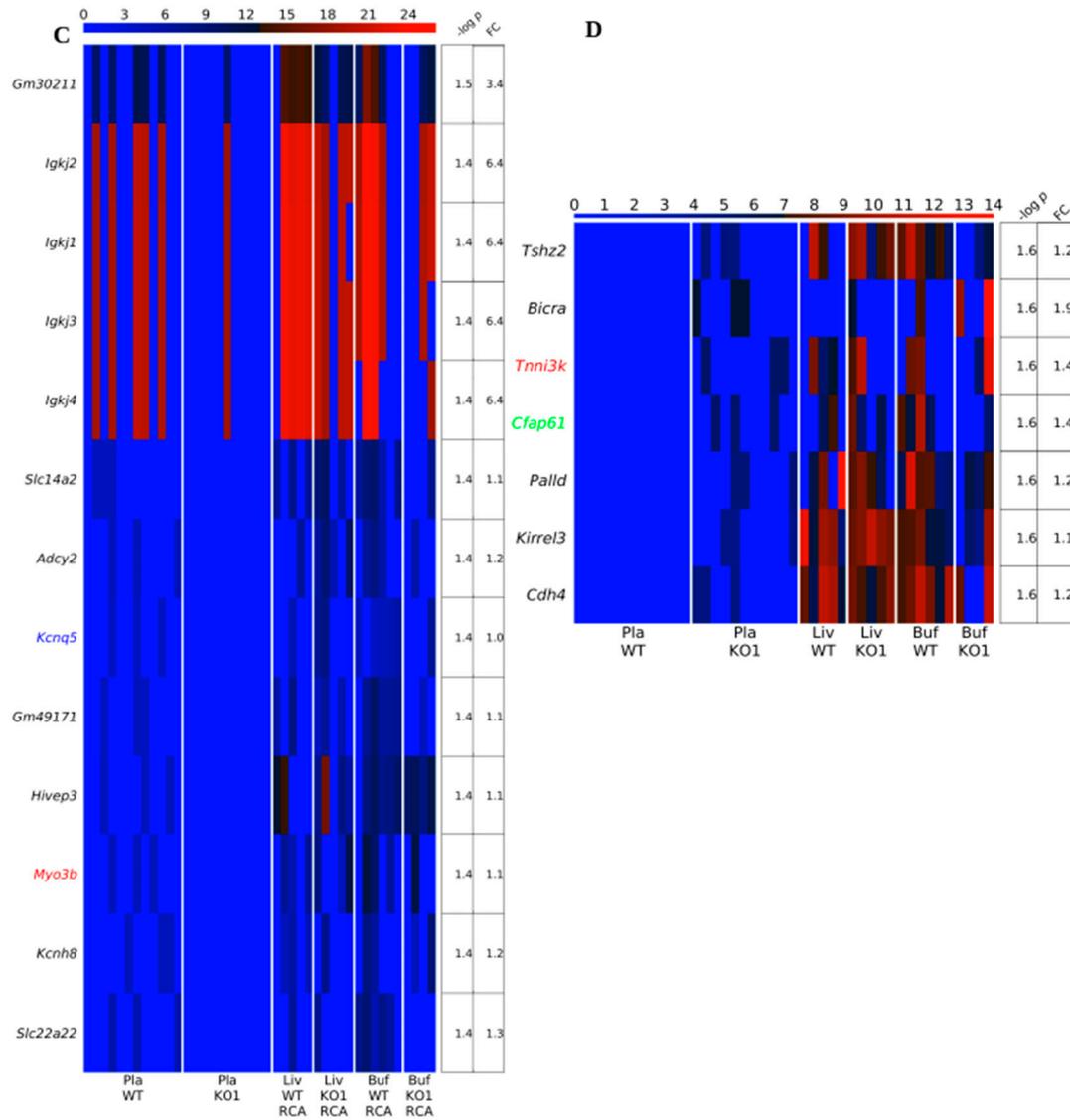


Figure 3. Cont.



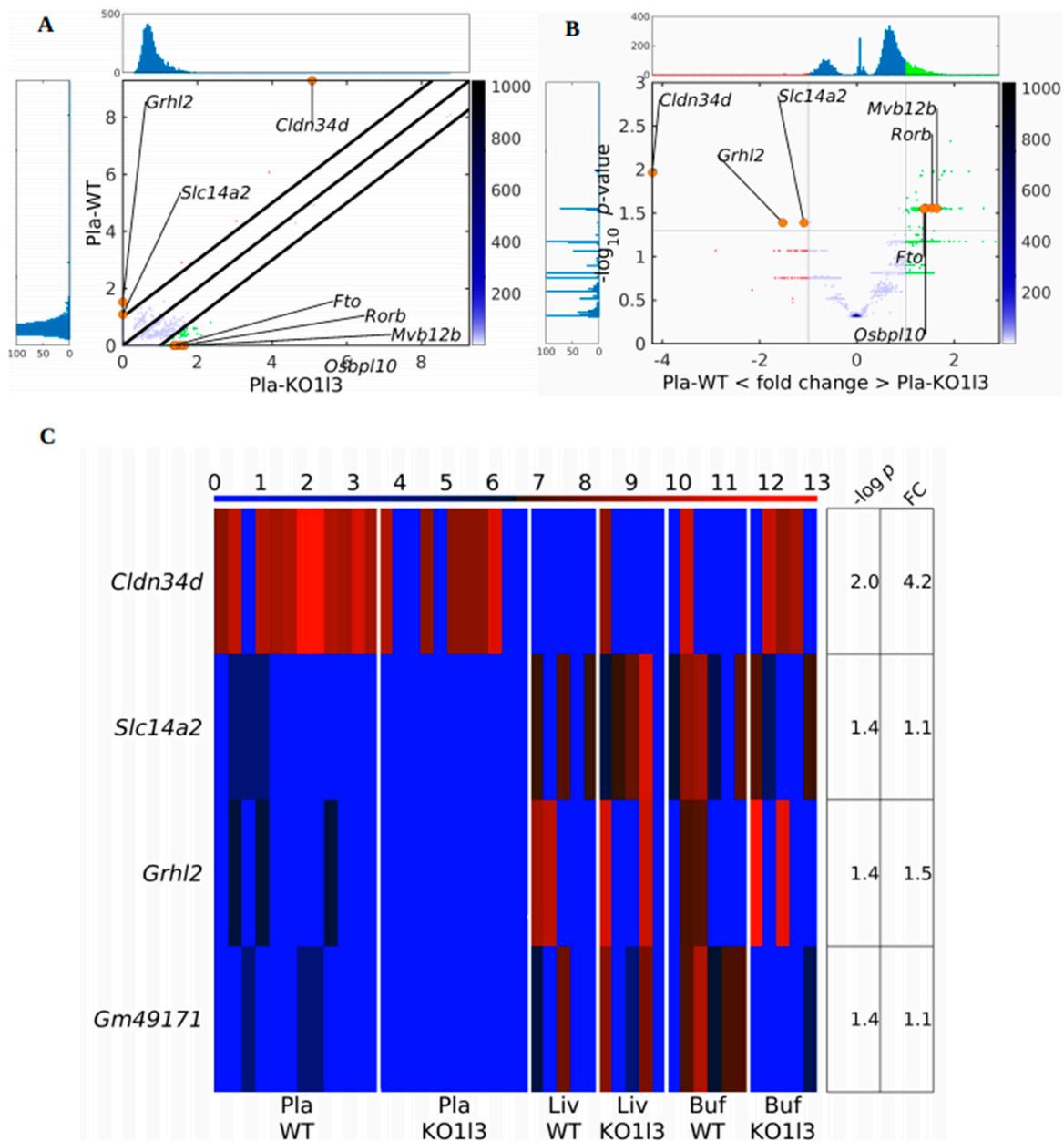
**Figure 3.** Scatter plot (A) and volcano plot (B) of the PpGCs in Pla-KO1 vs. Pla-WT samples. In (B), the vertical black lines are the boundaries of the  $\log_2(FC)$  1 thresholds for the eccDNA production levels between the groups. The horizontal black line is the 1.3-boundary of the  $p$ -value in  $-\log_{10}$  scale (corresponding to  $p$ -value = 0.05 in linear scale). PpGCs beyond these thresholds were inferred to be significantly distinct from control eccDNAs and denominated as DPpGCs. Their distributions are visualized with histograms. Red and green dots mark PpGCs with abundance below and above the FC threshold. Orange dots indicate selected DPpGCs. Heatmaps of the (C) 13 genes that led to up-DPpGCs in WT relative to Pla-KO1 and (D) 7 genes that led to up-DPpGCs in Pla-KO1 relative to WT in decreasing order of significance. The PpGCs in the liver samples, Liv-WT-RCA and Liv-KO1-RCA, and in the buffy coat samples, Buf-WT-RCA and Buf-KO1-RCA, are given in the heatmaps as a comparison for eccDNA production in the corresponding tissues. The color bars codify the value of PpGCs on the  $\log_2$  scale. A higher value corresponds to a redder color. The  $-\log_{10}(p$ -value) and the absolute value of the  $\log_2$  of the fold change (FC) in the DPpGCs are presented in a table to the right of the heatmaps. The gene names at the left side of the heatmaps are colored in red, green, blue, cyan, and magenta for genes associated with CFSs induced by aphidicolin, folic acid, BrdU, 5-azacytidine, and distamycinA, respectively.

**Table 1.** Enrichment of CFSs in DPpGCs between DNase KOs and controls in mouse plasma and human samples. The percentages are the ratios of the numbers of CFSs in genes multiplied by 100 and divided by the number of DPpGCs. The  $p$ -value of the statistical significance is calculated by the mid  $p$ -values of the hypergeometric distribution. The statistically significant cases are marked in boldface. WT wild type, HC healthy control.

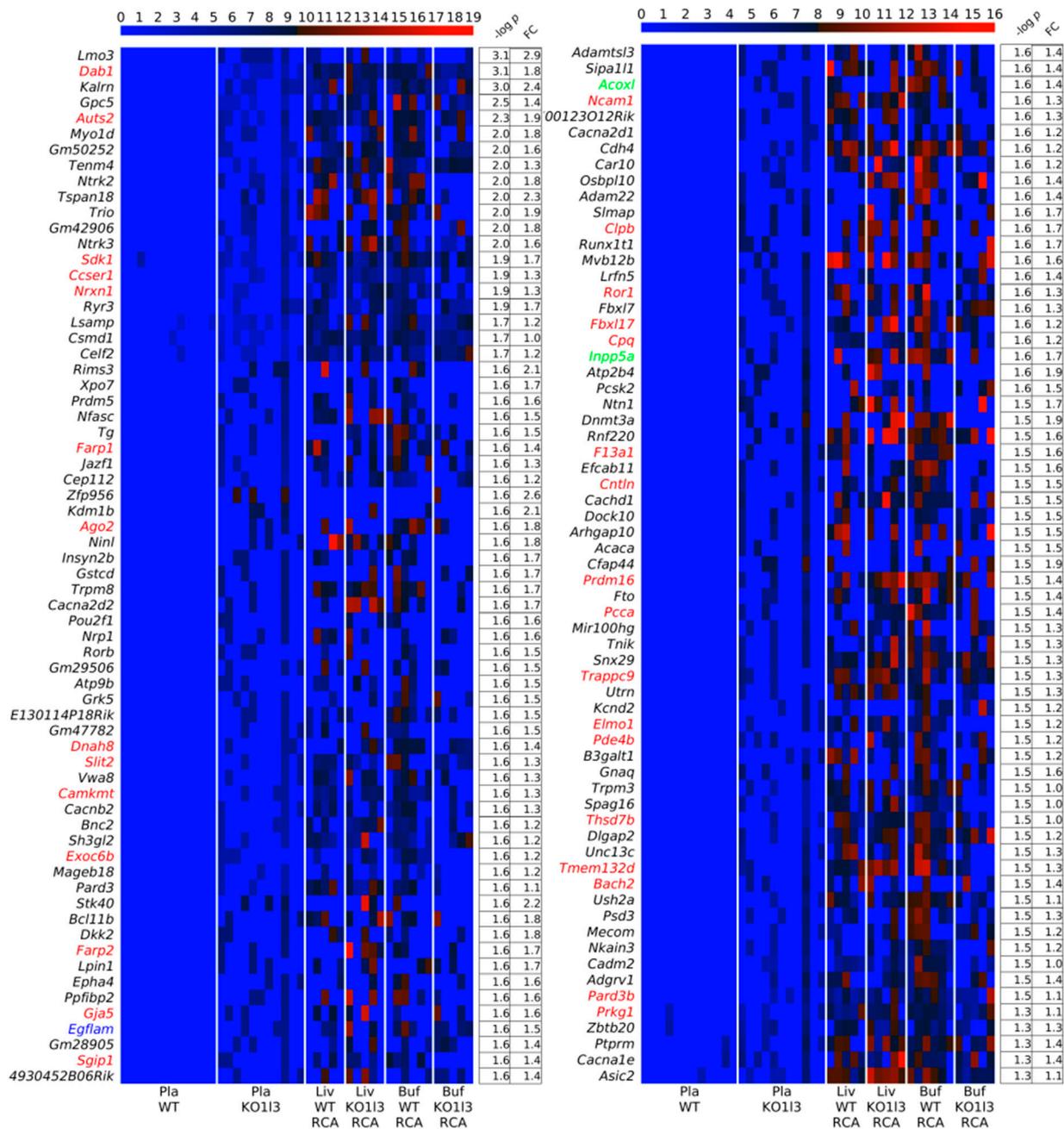
Species	DNase Mutation	Control	Up-DPpGC	Percentage	$p$ -Value
Mouse	KO1	WT	WT	15.38%	0.2189
Mouse	KO1	WT	KO1	28.57%	0.0730
Mouse	KO113	WT	WT	0%	0.6578
<b>Mouse</b>	<b>KO113</b>	<b>WT</b>	<b>KO113</b>	<b>25.95%</b>	<b><math>8.1695 \cdot 10^{-9}</math></b>
<b>Human</b>	<b>DNASE1L3</b>	<b>HC</b>	<b>KO113</b>	<b>20.97%</b>	<b><math>5.298 \cdot 10^{-11}</math></b>

### 3.4. Distinctive and Specific Genic cf-eccDNA Profile of *Dnase1l3* KO Compared to WT

In the plasma *Dnase1l3* KO vs. WT case, the traditional approaches, comparisons of numbers and lengths of eccDNAs, as in the case of *Dnase1* KO vs. WT, did not find significant differences between the sample groups using circulomics data. Additionally, in the comparison of Pla-KO1 vs. Pla-WT, we found only scarce difference with our gene-centric approach. However, to assess whether any of the genes in the mouse genome were particularly prone to form cf-eccDNA in the *Dnase1l3* KO group compared to the WT group, we calculated the PpGCs for the Pla-KO113 and Pla-WT samples. The paired scatter plot in Figure 4A shows differences in PpGCs in the Pla-KO113 and Pla-WT samples. For a  $\log_2$  fold change  $\geq 1$  ( $FC \geq 2$  on the linear scale) and a  $-\log_{10}(p\text{-value}) \geq 1.3$ , i.e., a significance level threshold  $\alpha_{\text{DPpGC}} = 0.05$ , we identified 131 up-DPpGCs in Pla-KO113 related to Pla-WT (Figures 4B and 5) and 4 up-DPpGCs in Pla-WT compared to Pla-KO113 (Figure 4B,C), namely, *Cldn34d*, *Slc14a2*, *Grhl2* and *Gm49171*. The statistically most significant up-DPpGC in Pla-KO113 is *Lmo3*, LIM domain only 3, which is predicted to enable ion binding activity, with a  $p$ -value of 0.00079 (Figure 5), while the top up-DPpGC in Pla-WT compared to Pla-KO113 is *Cldn34d*, claudin 34D, which is involved in bicellular tight junction assembly and cell adhesion, with a much higher  $p$ -value of 0.01080 (Figure 4C). Anyway, *Cldn34d* shows high PpGC values even in Pla-KO113 samples and thus is not specific for Pla-WT. Next, we checked whether the genic circulomics difference in plasma correlates with that in liver and buffy coat tissues. We chose to show in the heatmaps in Figures 4C and 5 the PpGCs in tissue samples from libraries obtained with the RCA method since using this method in tissues led to the mapping of a high number of eccDNAs, while the Tn method resulted in mapping very few eccDNA. The same heatmaps of PpGCs in liver and buffy coat obtained with the Tn method are represented in Supplementary Figures S2 and S3. Interestingly, for the genes up-producing cf-eccDNA in Pla-KO113 and Pla-WT, the PpGCs do not differentiate between the KO and the WT groups in both liver and buffy coat tissue. In conclusion, the genic cf-eccDNA profile of Pla-KO113 vs. Pla-WT is distinctive with a high number of genes; it is specific, with cf-eccDNA from these genes produced generally only in Pla-KO113 and none in Pla-WT; and it does not correlate with a difference from the same genes in liver and buffy coat tissues.



**Figure 4.** Scatter plot (A) and volcano plot (B) of the PpGCs in Pla-KO113 vs. Pla-WT samples. In (A), the parallel black lines are the boundaries of the  $\log_2(\text{FC})$  1 thresholds for the eccDNA production levels between the groups. In (B), the vertical black lines are the boundaries of the  $\log_2(\text{FC})$  1 thresholds for the eccDNA production levels between the groups. The horizontal black line is the 1.3-boundary of the  $p$ -value on the  $-\log_{10}$  scale (corresponding to  $p$ -value = 0.05 in linear scale). PpGCs beyond these thresholds were inferred to be significantly distinct from control eccDNAs and denominated as DPpGCs. Their distributions are visualized in histograms. Red and green dots mark PpGCs with abundance below and above the FC threshold. Orange dots indicate selected DPpGCs. (C) Heatmap of the four genes that led to up-DPpGCs in Pla-WT relative to Pla-KO113 in decreasing order of significance. The PpGCs in the liver samples, Liv-WT-RCA and Liv-KO113-RCA, and in the buffy coat samples, Buf-WT-RCA and Buf-KO113-RCA, are given in the heatmaps as a comparison for the eccDNA production in the corresponding tissues. The color bars codify the value of PpGCs on the  $\log_2$  scale. A higher value corresponds to a redder color. The  $-\log_{10}(p\text{-value})$  and the absolute value of the  $\log_2$  of the fold change (FC) of the DPpGCs are presented in a table to the right of the heatmap. The gene names to the left of the heatmaps are colored in red, green, blue, cyan, and magenta for genes associated with CFSs induced by aphidicolin, folic acid, BrdU, 5-azacytidine, and distamycinA, respectively.



**Figure 5.** Heatmaps of the 131 genes that led to up-DEPpGCs in Pla-KO113 compared to WT in decreasing order of significance. The PpGCs in the liver samples, Liv-WT-RCA and Liv-KO1-RCA, and in the buffy coat samples, Buf-WT-RCA and Buf-KO1-RCA, are given in the heatmaps as a comparison for the eccDNA production in the corresponding tissues. The color bars codify the value of PpGCs on the log<sub>2</sub> scale. A higher value corresponds to a redder color. The  $-\log_{10}(p\text{-value})$  and the absolute value of the log<sub>2</sub> of the fold change (FC) of the DEPpGCs are presented in a table to the right of the heatmaps. The gene names to the left of the heatmaps are colored in red, green, blue, cyan, and magenta for genes associated with chromosomal fragile sites (CFSs) induced by aphidicolin, folic acid, BrdU, 5-azacytidine, and distamycinA, respectively.

### 3.5. Association of up-DEPpGCs in Pla-KO113 with CFSs

We checked our list of up-DEPpGCs in Pla-KO113 for genes from chromosomal fragile sites (CFSs), characterized as common and rare, as genomically unstable regions, which are

hot-spots for deletions and other alterations. We found that disabled-1, *Dab1*, the second ranked one, has been characterized as a large CFS gene that is mapped within the FRA1B CFS and inactivated in multiple cancers [10,21], while *AUTS2* (rank 5) maps in FRA7J, *NRXN1* (rank 16) in FRA2D, *PRKG1* (127) in FRA10C [10], and *SDK1* (14) in FRA7B [22]. We checked the other up-DPpGCs in the HumCFS database [19], as of 6 October 2023, and found an additional 29 up-DPpGCs as genes associated with CFSs. Thus, 34 of the 131 up-DPpGCs in Pla-KO113, or 26%, are genes mapped to characterized CFSs. This percentage is similar to the 29% of up-DPpGCs in Pla-KO1. Anyway, while in Pla-KO1 the up-DPpGCs are not enriched in CFS genes, in PlaKO113, this enrichment is statistically significant with a *p*-value of  $8.1695 \cdot 10^{-9}$  (Table 1). The CFS-associated genes are listed in Table 2 together with their rank as up-DPpGCs and their association with a CFS in humans and the corresponding mouse chromosome, and additionally marked in Figure 5 with a color corresponding to the agent with which they have been discovered. Of the six genes on human chromosome 1, five map to mouse chromosome 4 and one to mouse chromosome 3. The three genes on human chromosome 8 map to mouse chromosome 15. The rest of the human CFC genes have a more disperse mapping to mouse chromosomes.

**Table 2.** Human homologs of the up-DPpGCs in Pla-KO113 compared to Pla-WT, their rank as up-DPpGCs, their belonging to a certain chromosomal fragile site (CFS) in humans and their corresponding mouse chromosome. In boldface are marked the cases with more stable human-to-mouse chromosome mapping.

Gene	Rank of DPpGC	CFS	Mouse Chromosome
<b>DAB1</b>	<b>2</b>	<b>FRA1B</b>	<b>4</b>
<i>AUTS2</i>	5	FRA7J	5
<i>SDK1</i>	14	FRA7B	5
<i>CCSER1</i>	15	FRA4E	6
<i>NRXN1</i>	16	FRA2D	17
<i>FARP1</i>	26	FRA13D	14
<b>AGO</b>	<b>31</b>	<b>FRA8D</b>	<b>15</b>
<i>DNAH</i>	45	FRA6H	17
<i>SLIT2</i>	46	FRA4D	5
<i>CAMKMT</i>	48	FRA2S	17
<i>EXOC6B</i>	52	FRA2E	6
<i>FARP2</i>	58	FRA2J	1
<i>GJA5</i>	62	FRA1F	3
<i>EGFLAM</i>	63	FRA5A	15
<b>SGIP1</b>	<b>65</b>	<b>FRA1L</b>	<b>4</b>
<i>ACOXL</i>	69	FRA2B	2
<i>NCAM1</i>	70	FRA11G	9
<i>CLPB</i>	78	FRA11H	7
<b>ROR1</b>	<b>82</b>	<b>FRA1L</b>	<b>4</b>
<i>FBXL17</i>	84	FRA5F	17
<b>CPQ</b>	<b>85</b>	<b>FRA8B</b>	<b>15</b>
<i>INPP5A</i>	86	FRA10A	7
<i>F13A1</i>	92	FRA6B	13
<i>CNTLN</i>	94	FRA9G	4
<b>PRDM16</b>	<b>100</b>	<b>FRA1A</b>	<b>4</b>
<i>PCCA</i>	102	FRA13D	14
<b>TRAPPC9</b>	<b>106</b>	<b>FRA8D</b>	<b>15</b>
<i>ELMO1</i>	109	FRA7C	13
<b>PDE4B</b>	<b>110</b>	<b>FRA1L</b>	<b>4</b>
<i>THSD7B</i>	115	FRA2F	1
<i>TMEM132D</i>	118	FRA12E	5
<i>BACH2</i>	119	FRA6G	4
<i>PAR3B</i>	126	FRA21	1
<i>PRKG1</i>	127	FRA10C	19

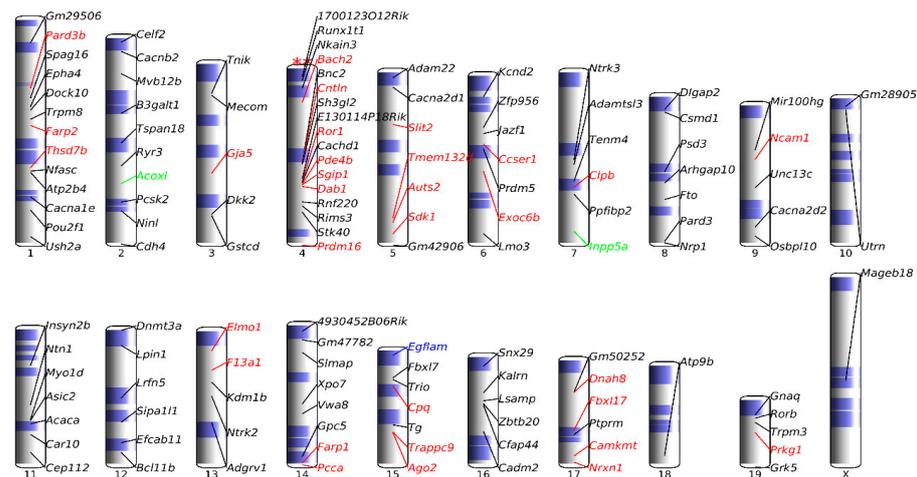
Next, we calculated the statistical enrichment of the CFS genes. We found that the genic cf-eccDNA profile of *Dnase1l3*<sup>-/-</sup> mice compared to WT mice is statistically significantly enriched in CFS genes (Table 1).

### 3.6. Association of up-DPpGCs in Pla-KO1l3 with SLE in the GWAS catalog

We made an intersection of the up-DPpGCs in Pla-KO1l3 with the genes for ‘systemic lupus erythematosus’ in the GWAS catalog and found three genes in common, namely, *JAZF1*, *DOCK10*, and *BACH2*. Gateva et al. [23] found the promoter region of *JAZF1*, juxtaposed with another zinc finger gene 1, as a confirmed SLE locus (rs849142). We found *JAZF1* to be the only gene common with the GWAS term ‘systemic lupus erythematosus and systemic sclerosis’. *DOCK10*, dedicator of cytokinesis 10, belongs to a cytokinesis protein family, whose members are guanosine nucleotide exchange factors for Rho GTPases. It has been confirmed as a SLE locus (rs11462616) in an Egyptian population [24]. Sheng et al. [25] suggested that overexpression of the transcriptional regulator *BACH2* represses Th9 cell differentiation by suppressing *IRF4* expression in SLE patients, and that it might be a potential target for SLE treatment. *BACH2*, BTB and CNC homology 1, basic leucine zipper transcription factor 2, is the only one of the three genes associated with a common CFS, FRA6G.

### 3.7. Chromosomal Landscaping of up-DPpGCs in Pla-KO1l3 Compared to Pla-WT

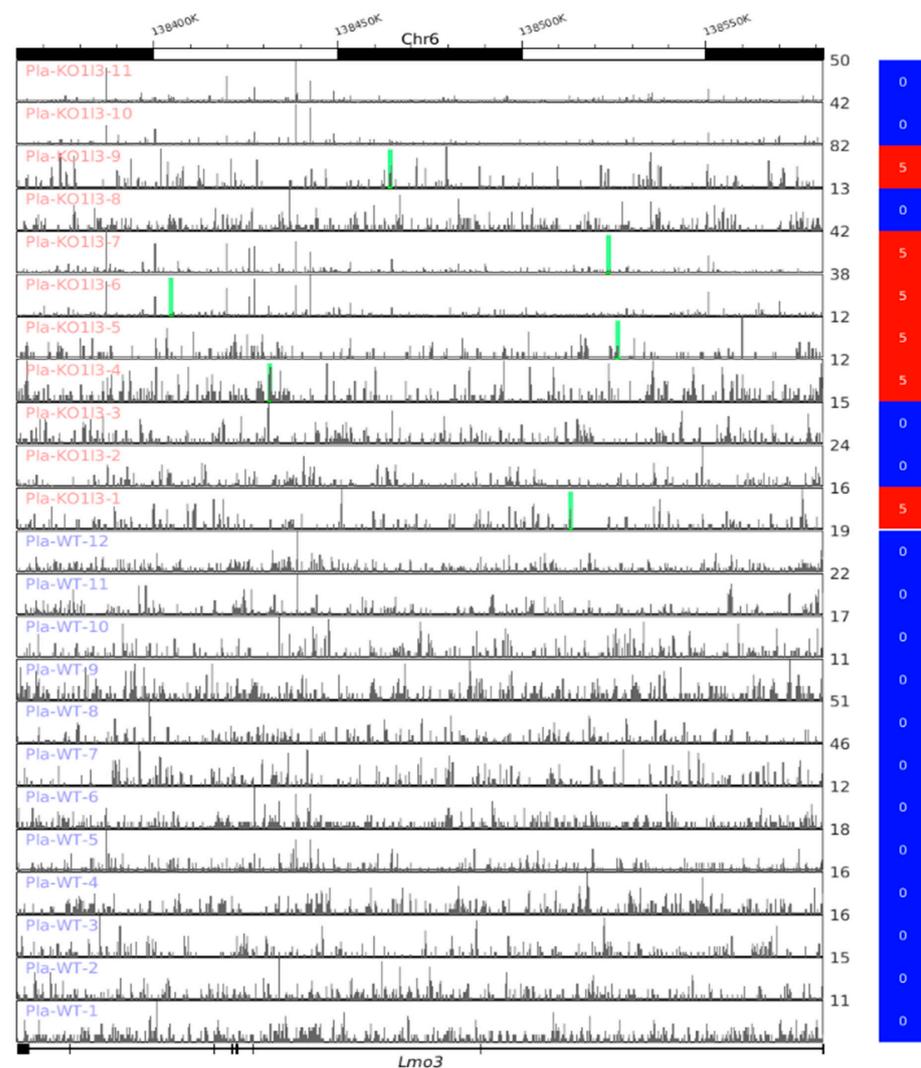
Our DifCir method revealed a high number of up-DPpGCs in *Dnase1l3* KO vs. WT plasma. To obtain a better insight in the distribution of the DPpGCs and the DPpGCs that are CFS genes across the chromosomes, we performed chromosomal landscaping of the distribution of these up-DPpGCs, which showed that chromosome 4 is statistically significantly enriched in DPpGCs in *Dnase1l3* KO mice (Figure 6). Additionally, we observed in the chromosomal landscape that all the CFS-related genes on chromosome 4 have been produced by aphidicolin. On chromosome 4, there seems to be a cluster of four CFS-associated genes (*Ror1*, *Pde4b*, *Sgip1*, and *Dab1*). The first three are from the human FRA1L CFS, and the last one is from the FRA1B CFS (Table 2). *Cachd1* belongs to the same cluster and is not a known CFS gene; anyway, its location between three FRA1L genes suggests that it might also be a FRA1L gene, and at the same time, *Dab1* does not exactly belong to the cluster. On chromosome 14, *Farp* and *Pcca* belong to FRA13D, and on chromosome 15, *Trappc9* and *Ago2* associate with FRA8D (Table 2).



**Figure 6.** Chromosomal landscape of genomic loci giving rise to statistically significant up-DPpGCs in Pla-KO1l3 compared to Pla-WT. The colored genes are associated with chromosomal fragile sites (CFSs): red, green, blue, cyan, and magenta are for genes associated with CFSs induced by aphidicolin, folic acid, BrdU, 5-azacytidine, and distamycinA, respectively. Chromosome Y has not been included since there were no DPpGCs excised from this chromosome. \*\* over chromosome 4 means that it has a significant enrichment of up-DPpGCs with  $0.001 \geq \text{mid-}p\text{-value} > 0.0001$ .

### 3.8. Visualization of up-DPpGCs in Pla-KO113, eccDNA Excising Loci and Coverage in All Plasma Samples

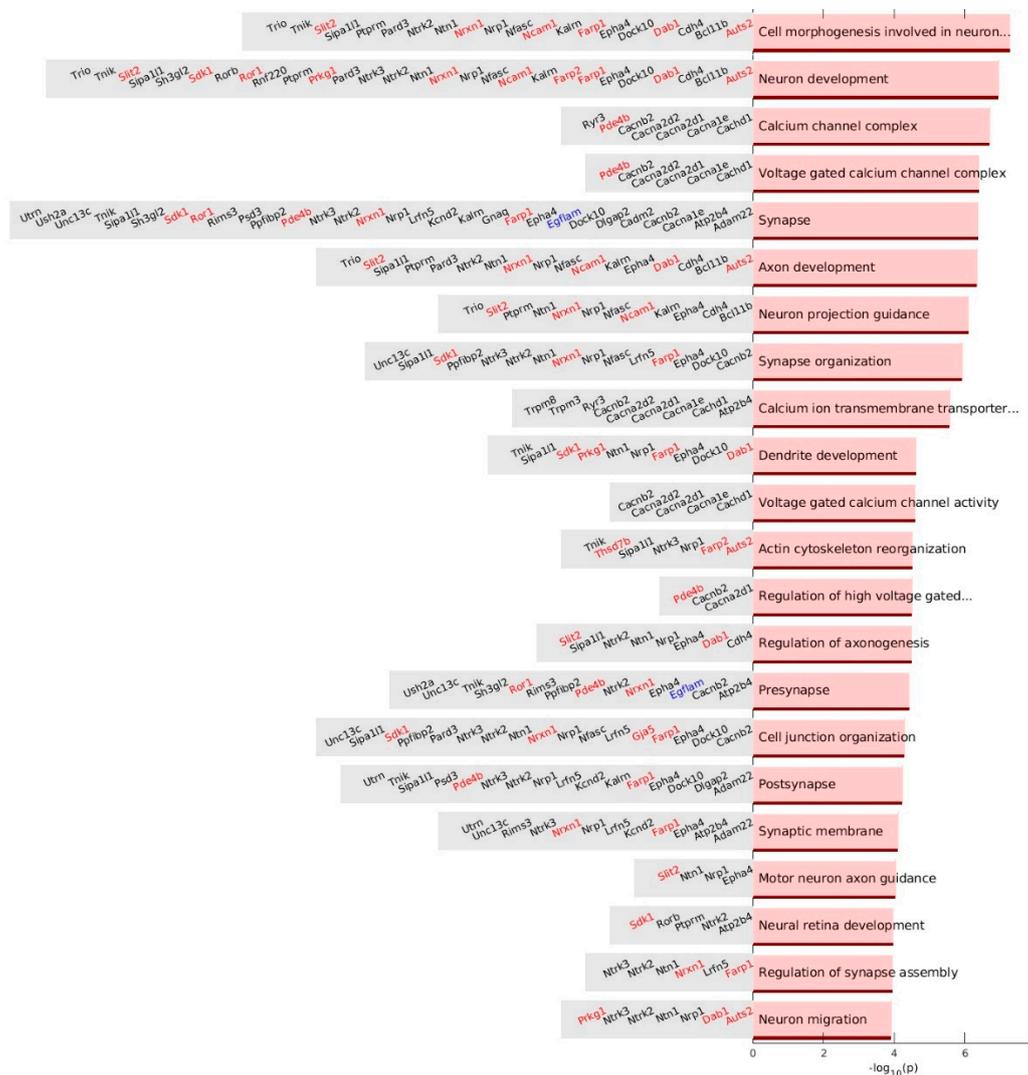
We explored the relation between the eccDNAs leading to the up-DPpGCs in Pla-KO113 and the coverage of the associated genes through depicting simultaneously the genomic coordinates of the gene from which the eccDNAs were excised, the coverage in that gene in all samples, and the value of the quantity of PpGCs for the gene and the sample. Figure 7 shows the top up-DPpGCs in Pla-KO113, *Lmo3*, illustrative in that the eccDNA forming the Pla-KO113 genic profile carries fragments of genes rather than whole genes, that these fragments do not have the same coordinates, and that most of the coverage, if not all, is not related to mapped eccDNA. In the case of *Lmo3*, only one unique eccDNA has been excised per sample in the Pla-KO113 samples where eccDNA has been mapped, i.e., samples 1, 4–7, and 9. The track plots of the next 10 up-DPpGCs in the ranking are shown in Supplementary Figures S5–S14. An example of more than one eccDNA sequence excised from a gene is *Auts2* (Supplementary Figure S7), where Pla-KO113 samples 1, 2, 4, 6 and 9 have three, one, one, one and eight eccDNAs, respectively, but the PpGC values corresponding to these samples are five, four, three, three and six, due to the different numbers of split reads mapping each unique eccDNA sequence.



**Figure 7.** Track plots of the loci the top-ranked up-DPpGC in Pla-KO113 compared to WT, *Lmo3*, and the corresponding gene coverage in all the plasma samples from the two groups. Each horizontal line represents the length of a gene. The green bars represent the excision loci of the eccDNA. The color bars to the right codify the value of PpGCs on the  $\log_2$  scale.

### 3.9. Gene Set Enrichment Analysis of up-DPPGCs in Pla-KO113

We performed a gene ontology (GO) enrichment analysis using the gene set enrichment analysis (GSEA) sets for systematic functional associations of the genes related to the up-DPPGCs. The top enriched ontology term of the up-DPPGCs in Pla-KO113 is ‘cell morphogenesis involved in neuron differentiation’ due to the detection of the up-DPPGCs in 20 genes: *Auts2*, *Bcl11b*, *Cdh4*, *Dab1*, *Dock10*, *Epha4*, *Farp1*, *Kalrn*, *Ncam1*, *Nfasc*, *Nrp1*, *Nrxn1*, *Ntn1*, *Ntrk2*, *Pard3*, *Ptpnm*, *Sipa11l*, *Slit2*, *Tnik*, and *Trio* (Figure 8). We checked whether some terms are enriched in CFS genes. The genes associated with neural-related GSEA terms have a high percentage of CFSs: ‘cell morphogenesis involved in neuron differentiation’ has 30% CFS genes and ‘neuron development’ has 36% CFS genes. Only two terms, ‘calcium ion transmembrane transporter activity’ and ‘voltage-gated calcium channel activity’, are represented by sets of genes that have no CFS-associated genes (Figure 8).



**Figure 8.** GSEA terms from the enrichment analysis of up-DPPGCs in Pla-KO113. **(Right)** Bar plots of the  $-\log_{10}(p\text{-value})$  of the significantly enriched terms. **(Left)** List of genes in the significant GSEA enrichment terms. The colored genes are associated with chromosomal fragile sites (CFSs): red, green, blue, cyan, and magenta are for genes associated with CFSs induced by aphidicolin, folic acid, BrdU, 5-azacytidine, and distamycinA, respectively.

Additionally, we performed a GO enrichment analysis separately for Molecular functions, Biological processes and Cellular components, as presented in Supplementary Figures S14–S16. The top term for Molecular functions is ‘calcium ion transmembrane

transporter activity', which does not have any CFS-associated gene. Interestingly, among the GO Molecular function terms are 'Rho guanyl-nucleotide exchange factor activity' (Supplementary Figure S14) with the genes *Dock10*, *Farp1*, *Farp2*, *Kalrn*, and *Trio*. An understanding of how guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) are regulated and interact with their GTPase targets is providing new opportunities for therapeutic development in immune-mediated inflammatory diseases [26]. The top term for Biological processes is 'regulation of axonogenesis', and among the terms is 'regulation of GTPase activity' with the genes *Epha4*, *Farp1*, *Farp2*, *Gnaq*, *Kalrn*, *Ntrk2*, *Ntrk3*, *Prkg1*, *Sipa11l1*, and *Slit2* (Supplementary Figure S15). A KALRN variant and risk allele rs1444766-G are associated with SLE. The PRKG1 rs7897633 variant has been previously identified as the top hit in patients of European ancestry with SLE and high IFN- $\alpha$  levels compared to those with low circulating IFN activity [27]. Slit2 glycoprotein has been found to be a potential biomarker for renal impairment in SLE patients [28].

The first statistically significantly enriched term for GO Cellular components is 'synapse' with gene members *Adgrv1*, *Asic2*, *Cadm2*, *Cep112*, *Dlgap2*, *Egflam*, *Epha4*, *Farp1*, *Kcnd2*, *Nrxn1*, *Ntrk2*, *Pde4b*, *Psd3*, *Rims3*, *Sdk1*, *Sipa11l1*, *Unc13c*, *Ush2a*, and *Utrn*, while the second statistically significantly enriched term is 'voltage-gated calcium channel complex' with members *Cacna1e*, *Cacna2d1*, *Cacna2d2*, *Cacnb2*, and *Pde4b* (Supplementary Figure S16). SLE T cells exhibit activation signaling anomalies, including a defective Ca<sup>(2+)</sup> response and increased nuclear factor of activated T-cells, NFAT, nuclear translocation. Nicolaou et al. [29] reported a defect in Kv1.3 trafficking to the immunological synapse (IS) of SLE T cells that might contribute to the Ca<sup>(2+)</sup> defect and proposed that Kv1.3 trafficking abnormalities contribute to the altered distribution of Ca<sup>(2+)</sup> signaling in SLE T cells, with these defects explaining in part the T cell hyperactivity and dysfunction documented in SLE patients.

### 3.10. Common up-DPpGCs in Pla-KO113 and Pla-SLE Patients

Next, we aimed to compare the gene content of the genic cf-eccDNA profile from the *Dnase113*<sup>-/-</sup> SLE mouse model and the genic cf-eccDNA profile of SLE patients with DNASE1L3 deficiency we identified in a previous work [14]. Common up-DPpGCs in Pla-KO113 with up-DPpGCs identified in human SLE plasma samples are *Rorb*, *Mvb12b*, *Osbp110*, *Fto*, *Tnik*, and *Arhgap10*. Nandakumar et al. [30] identified the multivesicular body protein MVB12b as a target for TANK-binding kinase 1 phosphorylation, which is essential for the sorting of DNA into extracellular vesicles (EVs) and stimulation of bystander cells to stimulate the cGAS-STING pathway. Thus, MVB12b was identified as a messenger responsible for packing and exporting DNA fragments in exosomes, and it was found that turning down the messenger, rather than spreading an immune signal, sends an alarm signal to the tissue needing protection, thus providing perspectives in relation to the treatment of autoimmune diseases such as SLE. TNIK/MAP4K7 (Traf2- and Nck-interacting kinase) functions as an activator of the Wnt signaling pathway, and its inhibition has been reported to have antitumor activity [31].

### 3.11. Associations of up-DPpGCs in Pla-SLE with CFSs

We checked the up-DPpGCs in SLE patients with DNASE1L3 deficiency for CFS genes and found that 56 out of the 267 genes, i.e., approximately 21%, are CFS genes (Figure 9 and Table 1). These CFS genes have been denoted with the color of the chemical agent used for their validation. Additionally, we checked for the statistically significant enrichment of the CFS-affected genes and found that the up-DPpGCs are indeed enriched in CFS genes (Table 1).

AP004372.1	FOXN3	RP11-1069G10.1
BARX2	UBE2E2	KCNC4
PTPN21	<b>BTBD9</b>	C3orf35
PARD6G	ARHGAP10	SERPINB2
<b>SNX8</b>	TRIM44	RP11-546K22.1
LINC01399	<b>RAI14</b>	ASB4
<b>ATP11B</b>	<b>SPRED2</b>	WDR25
SPECC1L	<b>SLC1A7</b>	LINC00887
<b>SMYD1</b>	<b>UNC5B</b>	<b>PSMA3</b>
<b>LY86</b>	<b>CFAP221</b>	NOCT
LINC01102	MMP7	<b>TMEM30A</b>
MFSD6	ZSWIM2	<b>IL18RAP</b>
SEPTIN7	RP11-92C4.3	RP11-392A14.11
<b>NFATC3</b>	CCDC168	POC5
LINC02325	<b>SLC35E4</b>	EFCAB14
LINC02325	TLN1	<b>DCUN1D1</b>
RORB	ALDH3A2	POLR2M
SPECC1L-ADORA2A	FAM102A	GNAI3
RHOA	<b>ZC3H8</b>	BEND5
<b>UNC13A</b>	SLC22A10	ADCYAP1R1
SHROOM2	RP11-67A1.2	RP11-432O12.1
AMBRA1	PIK3IP1-DT	<b>TRIM67</b>
GNAL	LINC01495	IFT57
AC026167.1	AL591893.1	<b>GCFC2</b>
ADRA1B	RP11-946L16.1	PSMB7
<b>ATXN7L1</b>	KRT7	OSBPL2
COL11A1	RP11-401O9.4	<b>FBXL5</b>
CHD7	INPP1	EMX2O5
PELI2	<b>SLC34A2</b>	RP11-126O1.4
<b>CRTAC1</b>	C3orf85	<b>GIT2</b>
MVB12B	<b>PDPN</b>	<b>SLC29A4</b>
DLEU2	RP11-314A5.7	<b>MALT1</b>
CPE	CNTN2	ELMOD1
MAP4K3	<b>KMT2D</b>	AC018462.2
<b>ST7</b>	RP11-415J8.3	CERK
<b>TET3</b>	RP11-390F4.6	PLXDC1
<b>KIF1B</b>	SCPEP1	<b>TUBA1C</b>
KCTD1	NPEPL1	C9orf147
TRIM2	NRL	RP11-1109F11.6
NPHP4	BDKRB2	HIPK2
NIBAN1	TNFSF11	PDLIM1
ANO10	AGBL4-AS1	IPMK
LINC01847	ITFG2-AS1	KIF24
<b>RALGAP2</b>	CPSF6	RDM1P5
RP11-559F19.4	RP11-744N12.3	RP11-26L21.1
PAX5	UAP1	<b>TESC</b>
SNAP25-AS1	DCK	LINC02056
GNA14	PIWIL4	FAT2
<b>SCN8A</b>	WAC	ALDH1L2
LRRC1	RP11-18A15.1	AC009505.4
<b>MAPRE2</b>	BID	CTD-2135J3.3
<b>PLEKHA7</b>	SRPRB	MORC1
CEP85L	LINC01119	ELOVL7
NCOA1	<b>MARCO</b>	<b>PRPF40A</b>
<b>MATN2</b>	LINC00881	RP11-210M15.2
LMNTD1	<b>CEP89</b>	RP11-403H13.1
ATG10	<b>SLC22A16</b>	GSPT1
ANGPT1	<b>RNF215</b>	ANKRD33B
LINC01234	KLHL22	MIR3936HG
VAV3	ABCA17P	RP11-156K13.1
MAP3K20	PHYH	ELP1
DENND1A	SPACA3	GNAZ
MAP7	<b>DYNC1L12</b>	HPCAL1
PVT1	PELATON	<b>ARRB1</b>
<b>EXT1</b>	KCNH3	GPR137C
OSBPL10	GUCY2EP	CLDN16
CLEC16A	C16orf72	RSPH14
<b>MAD1L1</b>	SERPINA5	ARLNC1
RP11-383A19.1	RP11-381N20.2	<b>ARHGAP17</b>
KLHL3	NCKAP1	<b>CUEDC1</b>
<b>SRPK2</b>	ONECUT1	JPH3
CHST11	LINC02801	EVC
ANKRD44	SLC30A4	NECTIN1
FRAS1	<b>ZNF664</b>	RAB10
RP11-622O11.6	SPATA18	<b>RADIL</b>
TEAD1	NT5C3A	NDST1
GALNT14	PCP4	CYTH1
RP11-537P24.1	LINC02577	PKNOX1
DNER	SUPT20H	TGFBR2
ATP8B4	NLRP14	B4GALT5
MAP2	RP11-519G16.3	IKZF1
IQGAP2	VIRMA	MYO7A
<b>ALMS1</b>	<b>ATP6V1E2</b>	<b>DENND4A</b>
HHAT	TGS1	RP11-799O21.4
SETBP1	CCNK	SYNJ2
FTO	IFI16	<b>USP48</b>
CREB5	CHD8	ATP6AP1L
CD96	LINC02316	RP11-380I10.4
TNIK	ALPK3	PINX1
XKR6		

**Figure 9.** List of the up-DPPGCs in cf-eccDNA from SLE patients with DNASE1L3 deficiency compared to healthy controls in decreasing order of statistical significance. The colored genes are associated with chromosomal fragile sites (CFSs) according to the HumCFS database [19]. The color of the gene names corresponds to the chemical agent used to induce the CFS: cyan: 5-azacytidine; red: aphidicolin; blue: BrdU; magenta: distamycinA; and green: folic acid [19].

#### 4. Discussion and Conclusions

Dnase1l3 deficiency leads to autoimmunity. Systemic lupus erythematosus (SLE) with a complete Dnase1l3 deficiency causing pediatric onset in humans is distinct from SLE in general in that it shows no gender bias, has increased incidence of lupus nephritis and

anti-neutrophil cytoplasmic bodies. Though associations between Dnase1 polymorphisms and SLE have been made in humans, causality in human systems has not been proved, with Dnase1 showing clinical improvement in some SLE patients but notably not in lupus nephritis patients [1]. Here, we found a distinctive genic eccDNA profile of *Dnase113*<sup>-/-</sup> mice compared to WT mice, with some eccDNAs being from the same genes as in the pediatric SLE patients with DNASE1L3 deficiency, namely *Rorb*, *Mob12b*, *Osbpl10*, *Fto*, *Tnik*, and *Arhgap10*. Additionally, we identified an existent but not so discriminating genic eccDNA profile of *Dnase1*<sup>-/-</sup> mice compared to WT. The difference in the genic eccDNA profiles might be related to their different associated SLE types.

In the case of the Pla-KO1 comparison, we found more up-DPpGCs in the wild type group, Pla-WT, than in the Pla-KO1 group. Contrarily, in the Pla-KO113 study, we found more up-DPpGCs in the Pla-KO113 group compared to the Pla-WT group. Interestingly, for both Pla-KO1 and Pla-KO113, we found higher percentages of CFSs in the KOs than in the WT, which in the Pla-KO113 case were statistically significant, pointing to a relationship between the genes excising eccDNA in the KOs and CFSs; a relationship that is not observed in the WT.

Extrachromosomal circular DNA in cancer (ecDNAs) frequently carry oncogenes and multi-drug resistance genes. A multitude of research suggests that a decrease in the number of ecDNAs as well as the amplified genes might result in a reversal of the cancer phenotype. For example, Yu et al. [32] provided a direct relationship between the loss of double minutes (DMs) and a decrease in ovarian cancer malignancy, showing that the drug gemcitabine decreases the number of DMs. Xu et al. [3] investigated the characteristics and dynamics of circulating eccDNAs in type 2 diabetes mellitus (T2DM) patients undergoing short-term intensive insulin therapy (SIIT) for inducing long-term glycemic remission. Genic cf-eccDNA profiles before and after treatment offer the possibility to follow the effect of drugs not only at the level of number of eccDNAs but also at the more precise level of actual genes and CFSs that are affected by the drug. Common CFSs are frequently affected in cancer and can be found in nearly all healthy individuals, opening possible avenues in the relation between CFSs and ecDNAs.

In conclusion, here, for the first time, we found commonalities between the genic cell-free circulomics profiles of a mouse model of SLE, *Dnase113*<sup>-/-</sup>, and of the human disease itself, monogenic SLE with DNASE1L3 deficiency. The common features are: firstly, both exhibit a genic cf-eccDNA profile with a high number of genes that is specific for the disease group; secondly, the two genic cf-eccDNA profiles share six common genes; and thirdly, both mouse model and human SLE profiles contain similar percentages of genes from CFSs. In comparison, the SLE mouse model, *Dnase1*<sup>-/-</sup>, shows a genic cf-eccDNA profile with very few genes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines12010080/s1>. Supplementary Materials file.

**Author Contributions:** Methodology, D.G. and M.J.A.-B.; Formal analysis, D.G., P.F.M., A.Z., and M.J.A.-B.; Investigation, D.G., P.F.M., A.Z. and M.J.A.-B.; Writing—original draft, D.G. and M.J.A.-B.; Supervision, M.J.A.-B.; Funding acquisition, M.J.A.-B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 899417, Ministerio de Ciencia e Innovación, Spain grant No. PID2020-119715 GB-I00/AEI/10.13039/501100011033, and Instituto de Salud Carlos III, Infrastructure of Precision Medicine associated with Science and Technology (IMPACT) of the Strategic Action in Health (iDATA-MP).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This study makes use of data generated by The Chinese University of Hong Kong (CUHK) Circulating Nucleic Acids Research Group, as reported by Sin et al. in JCI insight

(doi: 10.1172/jci.insight.156070). The data have been deposited in the European Genome-Phenome Archive hosted by the European Bioinformatics Institute (accession EGAS00001005873.).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Keyel, P.A. Dnases in health and disease. *Dev. Biol.* **2017**, *429*, 1–11. [[CrossRef](#)] [[PubMed](#)]
- Luo, X.; Zhang, L.; Cui, J.; An, Q.; Li, H.; Zhang, Z.; Sun, G.; Huang, W.; Li, Y.; Li, C.; et al. Small extrachromosomal circular DNAs as biomarkers for multi-cancer diagnosis and monitoring. *Clin. Transl. Med.* **2023**, *13*, e1393. [[CrossRef](#)] [[PubMed](#)]
- Xu, Z.; He, J.; Han, P.; Dai, P.; Lv, W.; Liu, N.; Liu, L.; Liu, L.; Pan, X.; Xiang, X.; et al. Plasma extrachromosomal circular DNA is a pathophysiological hallmark of short-term intensive insulin therapy for type 2 diabetes. *Clin. Transl. Med.* **2023**, *13*, e1437. [[CrossRef](#)] [[PubMed](#)]
- Di Vincenzo, F.; Yadid, Y.; Petito, V.; Emoli, V.; Masi, L.; Gerovska, D.; Araúzo-Bravo, M.J.; Gasbarrini, A.; Regenber, B.; Scaldaferrri, F. Circular and Circulating DNA in Inflammatory Bowel Disease: From Pathogenesis to Potential Molecular Therapies. *Cells* **2023**, *12*, 1953. [[CrossRef](#)] [[PubMed](#)]
- Gerovska, D.; Noer, J.B.; Qin, Y.; Ain, Q.; Januzi, D.; Schwab, M.; Witte, O.W.; Araúzo-Bravo, M.J.; Kretz, A. A distinct circular DNA profile intersects with proteome changes in the genotoxic stress-related hSOD1G93A model of ALS. *Cell Biosci.* **2023**, *13*, 170. [[CrossRef](#)] [[PubMed](#)]
- Yang, L.; Wang, M.; Hu, X.; Yuan, L.; Chen, S.; Peng, S.; Yang, P.; Yang, Z.; Bao, G.; He, X. EccDNA-oriented ITGB7 expression in breast cancer. *Ann. Transl. Med.* **2022**, *10*, 1344. [[CrossRef](#)] [[PubMed](#)]
- Sun, Z.; Ji, N.; Zhao, R.; Liang, J.; Jiang, J.; Tian, H. Extrachromosomal circular DNAs are common and functional in esophageal squamous cell carcinoma. *Ann. Transl. Med.* **2021**, *9*, 1464. [[CrossRef](#)]
- Sin, S.T.; Deng, J.; Ji, L.; Yukawa, M.; Chan, R.W.; Volpi, S.; Vaglio, A.; Fenaroli, P.; Bocca, P.; Cheng, S.H.; et al. Effects of nucleases on cell-free extrachromosomal circular DNA. *JCI Insight* **2022**, *7*, e156070. [[CrossRef](#)]
- Benitez, A.; Sebald, M.; Kanagaraj, R.; Rodrigo-Brenni, M.C.; Chan, Y.W.; Liang, C.C.; West, S.C. GEN1 promotes common fragile site expression. *Cell Rep.* **2023**, *42*, 112062. [[CrossRef](#)]
- Smith, D.I.; Zhu, Y.; McAvoy, S.; Kuhn, R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett.* **2006**, *232*, 48–57. [[CrossRef](#)]
- Helmrich, A.; Stout-Weider, K.; Hermann, K.; Schrock, E.; Heiden, T. Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res.* **2006**, *16*, 1222–1230. [[CrossRef](#)]
- Georgakilas, A.G.; Tsantoulis, P.; Kotsinas, A.; Michalopoulos, I.; Townsend, P.; Gorgoulis, V.G. Are common fragile sites merely structural domains or highly organized "functional" units susceptible to oncogenic stress? *Cell. Mol. Life Sci.* **2014**, *71*, 4519–4544. [[CrossRef](#)] [[PubMed](#)]
- Ji, F.; Zhu, X.; Liao, H.; Ouyang, L.; Huang, Y.; Syeda, M.Z.; Ying, S. New Era of Mapping and Understanding Common Fragile Sites: An Updated Review on Origin of Chromosome Fragility. *Front. Genet.* **2022**, *13*, 906957. [[CrossRef](#)] [[PubMed](#)]
- Gerovska, D.; Araúzo-Bravo, M.J. Systemic Lupus Erythematosus Patients with DNASE1L3-Deficiency Have a Distinctive and Specific Genic Circular DNA Profile in Plasma. *Cells* **2023**, *12*, 1061. [[CrossRef](#)] [[PubMed](#)]
- Gerovska, D.; Araúzo-Bravo, M.J. Skeletal Muscles of Sedentary and Physically Active Aged People Have Distinctive Genic Extrachromosomal Circular DNA Profiles. *Int. J. Mol. Sci.* **2023**, *24*, 2736. [[CrossRef](#)] [[PubMed](#)]
- Sin, S.T.K.; Jiang, P.; Deng, J.; Ji, L.; Cheng, S.H.; Dutta, A.; Leung, T.Y.; Chan, K.C.A.; Chiu, R.W.K.; Lo, Y.M.D. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1658–1665. [[CrossRef](#)] [[PubMed](#)]
- Shibata, Y.; Kumar, P.; Layer, R.; Willcox, S.; Gagan, J.R.; Griffith, J.D.; Dutta, A. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **2012**, *336*, 82–86. [[CrossRef](#)]
- Kumar, P.; Kiran, S.; Saha, S.; Su, Z.; Paulsen, T.; Chatrath, A.; Shibata, Y.; Shibata, E.; Dutta, A. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci. Adv.* **2020**, *6*, eaba2489. [[CrossRef](#)]
- Kumar, R.; Nagpal, G.; Kumar, V.; Usmani, S.S.; Agrawal, P.; Raghava, G.P.S. HumCFS: A database of fragile sites in human chromosomes. *BMC Genom.* **2019**, *19* (Suppl. S9), 985. [[CrossRef](#)]
- Uribe, M.L.; Dahlhoff, M.; Batra, R.N.; Nataraj, N.B.; Haga, Y.; Drago-Garcia, D.; Marrocco, I.; Sekar, A.; Ghosh, S.; Vaknin, I.; et al. TSHZ2 is an EGF-regulated tumor suppressor that binds to the cytokinesis regulator PRC1 and inhibits metastasis. *Sci. Signal.* **2021**, *14*, eabe6156. [[CrossRef](#)]
- McAvoy, S.; Zhu, Y.; Perez, D.S.; James, C.D.; Smith, D.I. Disabled-1 is a large common fragile site gene, inactivated in multiple cancers. *Genes. Chromosomes Cancer* **2008**, *47*, 165–174. [[CrossRef](#)]
- Bosco, N.; Pelliccia, F.; Rocchi, A. Characterization of FRA7B, a human common fragile site mapped at the 7p chromosome terminal region. *Cancer Genet. Cytogenet.* **2010**, *202*, 47–52. [[CrossRef](#)] [[PubMed](#)]
- Gateva, V.; Sandling, J.K.; Hom, G.; Taylor, K.E.; Chung, S.A.; Sun, X.; Ortmann, W.; Kosoy, R.; Ferreira, R.C.; Nordmark, G.; et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.* **2009**, *41*, 1228–1233. [[CrossRef](#)] [[PubMed](#)]

24. Elghzaly, A.A.; Sun, C.; Looger, L.L.; Hirose, M.; Salama, M.; Khalil, N.M.; Behiry, M.E.; Hegazy, M.T.; Hussein, M.A.; Salem, M.N.; et al. Genome-wide association study for systemic lupus erythematosus in an egyptian population. *Front. Genet.* **2022**, *13*, 948505. [[CrossRef](#)] [[PubMed](#)]
25. Sheng, Y.; Zhang, J.; Li, K.; Wang, H.; Wang, W.; Wen, L.; Gao, J.; Tang, X.; Tang, H.; Huang, H.; et al. Bach2 overexpression represses Th9 cell differentiation by suppressing IRF4 expression in systemic lupus erythematosus. *FEBS Open Bio.* **2021**, *11*, 395–403. [[CrossRef](#)] [[PubMed](#)]
26. Reedquist, K.A.; Tak, P.P. Signal transduction pathways in chronic inflammatory autoimmune disease: Small GTPases. *Open Rheumatol. J.* **2012**, *6*, 259–272. [[CrossRef](#)] [[PubMed](#)]
27. Fernandez Ruiz, R.; Shum, J.; Van Buren, K.; Niewold, T. Functional Effects of a Lupus-associated PRKG1 Variant on the RhoA-ROCK Pathway and Response to Type I Interferon [abstract]. *Arthritis Rheumatol.* **2021**, *73* (Suppl. S9), 1494. Available online: <https://acrabstracts.org/abstract/functional-effects-of-a-lupus-associated-prkg1-variant-on-the-rhoa-rock-pathway-and-response-to-type-i-interferon/> (accessed on 21 November 2023).
28. Zhang, Y.; Hu, L.; Li, X.; Chen, L.; Yang, X. Slit2 is a potential biomarker for renal impairment in systemic lupus erythematosus. *Clin. Exp. Med.* **2021**, *21*, 63–71. [[CrossRef](#)]
29. Nicolaou, S.A.; Neumeier, L.; Takimoto, K.; Lee, S.M.; Duncan, H.J.; Kant, S.K.; Mongey, A.B.; Filipovich, A.H.; Conforti, L. Differential calcium signaling and Kv1.3 trafficking to the immunological synapse in systemic lupus erythematosus. *Cell Calcium* **2010**, *47*, 19–28. [[CrossRef](#)]
30. Nandakumar, R.; Tschismarov, R.; Meissner, F.; Prabakaran, T.; Krissanaprasit, A.; Farahani, E.; Zhang, B.C.; Assil, S.; Martin, A.; Bertrams, W.; et al. Intracellular bacteria engage a STING-TBK1-MVB12b pathway to enable paracrine cGAS-STING signalling. *Nat. Microbiol.* **2019**, *4*, 701–713. [[CrossRef](#)]
31. Torres-Ayuso, P.; An, E.; Nyswaner, K.M.; Bensen, R.C.; Ritt, D.A.; Specht, S.I.; Das, S.; Andresson, T.; Cachau, R.E.; Liang, R.J.; et al. TNIK Is a Therapeutic Target in Lung Squamous Cell Carcinoma and Regulates FAK Activation through Merlin. *Cancer Discov.* **2021**, *11*, 1411–1423. [[CrossRef](#)] [[PubMed](#)]
32. Yu, L.; Zhao, Y.; Quan, C.; Ji, W.; Zhu, J.; Huang, Y.; Guan, R.; Sun, D.; Jin, Y.; Meng, X.; et al. Gemcitabine eliminates double minute chromosomes from human ovarian cancer cells. *PLoS ONE* **2013**, *8*, e71988. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.