



# Whole-Genome Sequencing of Vero E6 (VERO C1008) and Comparative Analysis of Four Vero Cell Sublines

Kazuhiro Konishi<sup>1</sup>, Toshiyuki Yamaji<sup>2</sup>, Chisato Sakuma<sup>2†</sup>, Fumio Kasai<sup>3</sup>, Toshinori Endo<sup>4</sup>, Arihiro Kohara<sup>3</sup>, Kentaro Hanada<sup>2\*†</sup> and Naoki Osada<sup>4\*</sup>

<sup>1</sup>Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan, <sup>2</sup>Department of Biochemistry and Cell Biology, National Institute of Infectious Diseases, Tokyo, Japan, <sup>3</sup>Laboratory of Cell Cultures, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan, <sup>4</sup>Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan

## OPEN ACCESS

### Edited by:

Henry Hq Heng,  
Wayne State University, United States

### Reviewed by:

Naazneen Khan,  
University of Kentucky, United States  
Christine Ye,  
University of Michigan, United States

### \*Correspondence:

Kentaro Hanada  
hanak@nih.go.jp  
Naoki Osada  
nosada@ist.hokudai.ac.jp

### †Present address:

Chisato Sakuma,  
Department of Bacteriology 1,  
National Institute of Infectious  
Diseases, Tokyo, Japan  
Kentaro Hanada,  
Department of Quality Assurance,  
Radiation Safety, and Information  
Management, National Institute of  
Infectious Diseases, Tokyo, Japan

### Specialty section:

This article was submitted to  
Human and Medical Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 25 October 2021

Accepted: 31 January 2022

Published: 22 March 2022

### Citation:

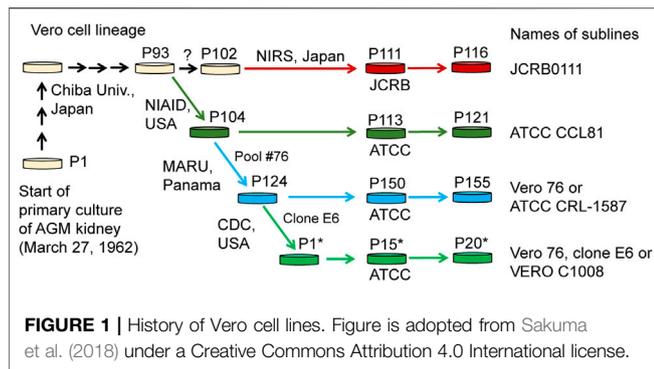
Konishi K, Yamaji T, Sakuma C,  
Kasai F, Endo T, Kohara A, Hanada K  
and Osada N (2022) Whole-Genome  
Sequencing of Vero E6 (VERO C1008)  
and Comparative Analysis of Four Vero  
Cell Sublines.  
Front. Genet. 13:801382.  
doi: 10.3389/fgene.2022.801382

The Vero cell line is an immortalized cell line established from kidney epithelial cells of the African green monkey. A variety of Vero sublines have been developed and can be classified into four major cell lineages. In this study, we determined the whole-genome sequence of Vero E6 (VERO C1008), which is one of the most widely used cell lines for the proliferation and isolation of severe acute respiratory syndrome coronaviruses (SARS-CoVs), and performed comparative analysis among Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6. Analysis of the copy number changes and loss of heterozygosity revealed that these four sublines share a large deletion and loss of heterozygosity on chromosome 12, which harbors type I interferon and *CDKN2* gene clusters. We identified a substantial number of genetic differences among the sublines including single nucleotide variants, indels, and copy number variations. The spectrum of single nucleotide variants indicated a close genetic relationship between Vero JCRB0111 and Vero CCL-81, and between Vero 76 and Vero E6, and a considerable genetic gap between the former two and the latter two lines. In contrast, we confirmed the pattern of genomic integration sites of simian endogenous retroviral sequences, which was consistent among the sublines. We identified subline-specific/enriched loss of function and missense variants, which potentially contribute to the differences in response to viral infection among the Vero sublines. In particular, we identified four genes (*IL1RAP*, *TRIM25*, *RB1CC1*, and *ATG2A*) that contained missense variants specific or enriched in Vero E6. In addition, we found that V739I variants of ACE2, which functions as the receptor for SARS-CoVs, were heterozygous in Vero JCRB0111, Vero CCL-81, and Vero 76; however, Vero E6 harbored only the allele with isoleucine, resulting from the loss of one of the X chromosomes.

**Keywords:** Vero cells, genome sequencing, Vero E6, cell lines, cell substrate

## INTRODUCTION

Cell lines established from mammalian tissues are often used for virus isolation and culture as well as for vaccine production. One of the cell lines frequently used for these purposes is the Vero cell line, which is an immortalized cell line established from kidney epithelial cells of an African green monkey (*Chlorocebus sabaues*, AGM) by Yoshihiro Yasumura at Chiba University in 1962. After the cell line was established, it was brought to the National Institute of Allergy and Infectious Diseases in the US, the American Type Culture Collection (ATCC), and the Japanese Collection of Research



Bioresources (JCRB) cell bank (**Figure 1**). Various sublines have been established through passaging (Earley and Johnson, 1988; Mizusawa et al., 1988; Terasima et al., 1988), which have slightly different properties from one another. For example, Vero CCL-81 (ATCC CCL-81), rather than Vero E6 (a.k.a. VERO C1008), is capable of more propagating Japanese encephalitis virus under prolonged culture conditions (Saito et al., 2020). Likewise, the subline Vero E6 likely propagates severe acute respiratory syndrome coronavirus (SARS-CoV) 2 more efficiently, compared with other Vero sublines (Harcourt et al., 2020; Matsuyama et al., 2020) and therefore Vero E6 and its derivatives have been widely used as the sublines for host cells in SARS-CoV-2 research (Basile et al., 2020; Ogando et al., 2020). However, the genetic factors that contribute to these phenotypic differences are largely unknown.

Whole-genome sequencing analysis of one of the sublines, Vero JCRB0111 was performed by Osada et al. (Osada et al., 2014) and an approximate 9-Mbp homozygous deletion on chromosome 12 was identified. The deletion was found to contain a cluster of type 1 interferon (*IFN-1*) genes that act as viral suppressors, as well as *CDKN2A* and *CDKN2B*, which are involved in the cell cycle control. These results indicate potential factors that resulted in the immortalization of Vero cells and may partly explain why viruses can readily multiply in these cells. Furthermore, whole-genome resequencing analysis of additional sublines, Vero CCL-81 and Vero 76 identified numerous genetic variations among the sublines as well as integrations of Simian endogenous retroviral sequences (SERVs) (Sakuma et al., 2018); however, the effects of the variants at the nucleotide level were not evaluated. More recently, Sene et al. reported the haplotype-resolved genome assembly of the Vero CCL-81 subline (Sene et al., 2021). Hundreds of protein-coding genes, including *ACE2*, were identified as losing function. AGM is susceptible to SARS-CoV-2 under laboratory conditions (Woolsey et al., 2021). Although the intrinsic function of *ACE2* is an angiotensin-converting, it is also known as the host cell receptor for SARS-CoV and SARS-CoV-2. Overall, these studies have provided information for quality control and have produced novel engineered sublines, which will accelerate the development of vaccine manufacturing platforms.

In the present study, we performed genome sequencing of Vero E6, which was established from a clone isolated from Vero

76 cells in earlier stage and conducted a comparative genome analysis among four major Vero cell sublines, currently available from public cell banks: Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6 (**Figure 1**). The identification of genetic variants specific or enriched in particular sublines will contribute to the elucidation of factors responsible for the phenotypic differences among the sublines.

## MATERIALS AND METHODS

### Data Preparation

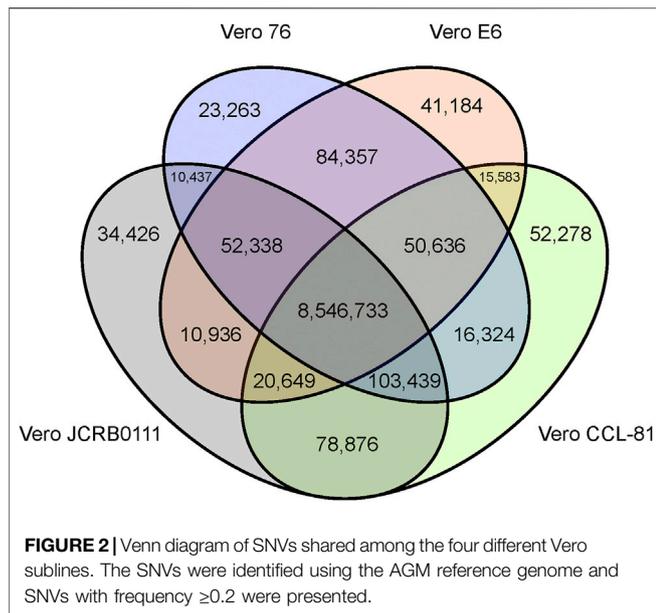
Vero E6 cells were obtained from ATCC (ATCC-CRL-1586). The short-read sequences of Vero JCRB0111, Vero CCL-81, and Vero 76 were downloaded from a public database (DDBJ: PRJDB2865). Paired-end sequences of Vero E6 were determined using an Illumina HiSeq 2500. The reads were mapped to the AGM reference genome (NCBI: GCF\_000409795.2) using the BWA-mem algorithm (Li and Durbin, 2009). The single nucleotide variants (SNVs) and short indels (<50 bp) were called using VarScan 2 software (Koboldt et al., 2012). The thresholds for SNV detection were:  $\geq 13$  coverage,  $\geq 2$  mutation count,  $\geq 15$  average base quality. The minimum coverage was selected to cover at least 95% of the genome for each sample. The strand filter was applied to exclude the sites where  $\geq 90\%$  of the reads mapped to one strand. Only bi-allelic SNVs were considered in this study. Manta was used to identify large-scale structural variations (Chen et al., 2015).

The GTF format file downloaded from the Ensembl database was used for gene annotation (Chlorocebus\_sabaeus.ChlSab1.1.86.gtf) and snpEFF software was used for annotating the effect of each genetic variant (Cingolani et al., 2012). The assignment for the gene function categories was performed using DAVID (Jiao et al., 2012). For evaluating the impact of missense variants, we used PROVEAN, SIFT, and PANTHER-PSEP (Choi et al., 2012; Choi and Chan, 2015; Tang and Thomas, 2016). The RNA-seq experiment data of Vero E6 Tmprss2+ (SRR13091741–SRR13091746) and COS-7 cells (SRR1919325–SRR1919327) were downloaded from the public database (Zhang et al., 2021). The RNA-seq reads were mapped to the reference genome using HISAT2 using the default parameters (Kim et al., 2019).

We defined subline-specific/enriched SNVs as variants with a significantly higher frequency compared with the other sublines. For each SNV, we counted the number of reads supporting the reference and alternative alleles and performed Tukey's test (Tukey, 1949). SNVs with  $p < 0.05$  were considered subline-specific/enriched SNVs after controlling for multiple-testing using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995).

### Copy Number Variations and Loss of Heterozygosity

We used Control-FREEC software to identify copy number variations (CNVs) and loss of heterozygosity (LOH) (Boeva et al., 2012). A window size of 50 kbp and a step size of



10 kbp were selected. To quantitate the genetic differentiation between sublines, we computed  $f_2$  statistics between all pairwise sublines.  $F_2$  was calculated as the sum of the squares of allele frequency differences between two sublines, divided by the total number of SNVs. The  $f_2$  values were used as a genetic distance between two sublines and a neighbor-joining tree was constructed using MEGA X (Knyaz et al., 2018).

### Genomic PCR for SVL27b, a SERV-Variation Locus on Chromosome 27

PCR was performed as described previously (Sakuma et al., 2018). Genomic DNA (30 ng) was used as the template. The primers specified below were synthesized (Eurofins Genomics Inc., Tokyo, Japan): Fw1: 5'-GGAACACCTGAAGATCTATGTGTC TA-3', Rv1: 5'-ATCAAATTCCTCTCTTCACATCTTCT-3', Fw2: 5'-GGACATATTGTTATAAAAGTTCATGG-3', Rv2: 5'-GAAACTATACCTATGATTTTGCCATAG-3'

## RESULTS AND DISCUSSION

### Genomic Features of the Vero Sublines

We obtained short reads from Vero E6 and mapped them to the AGM reference genome. To compare the genomic features of the Vero sublines, we reanalyzed previously published genomic data (Vero JCRB0111, Vero CCL-81, and Vero 76). In total, we obtained 74-, 32-, 43-, and 78-fold coverage data of the Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6 genomes, respectively. A 20% frequency cut-off yielded 9,141,259 SNVs, 453,797 short insertions, and 547,598 short deletions. As expected, 96% of these variants were shared among all sublines. The number of SNVs shared among sublines is summarized in **Figure 2**.

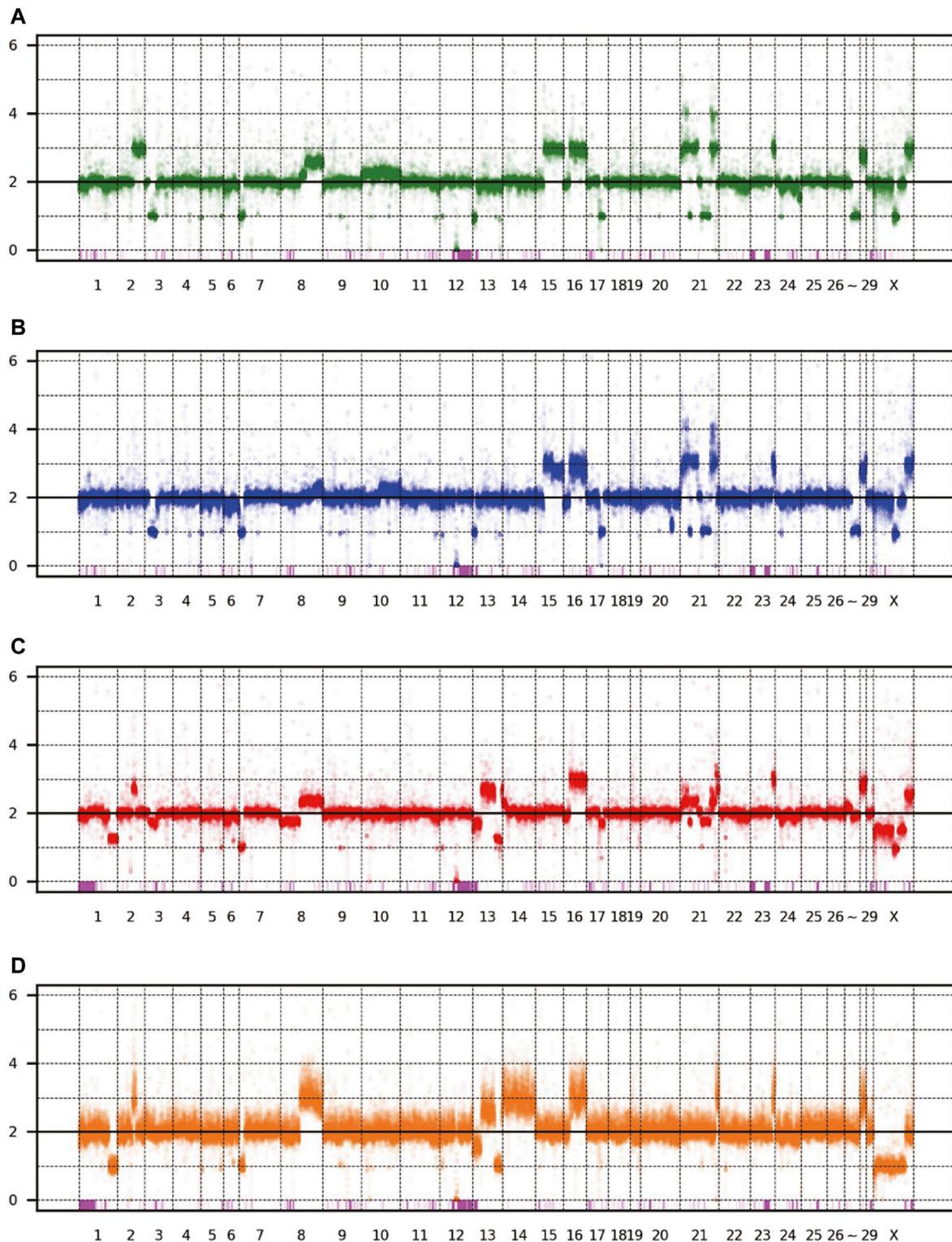
The patterns of CNV and LOH are shown in **Figure 3**, and the estimated copy numbers for each chromosome interval are provided in **Supplementary Table S1**. Overall, the patterns appeared similar among the four sublines, including the 9-Mbp deletion on chromosome 12. However, there were several marked differences; for example, a large part of chromosome 15 had three copies in Vero JCRB0111 and Vero CCL-81 but the region was normal with respect to copy number in Vero 76 and Vero E6. Interestingly, chromosome 21 is highly rearranged in Vero JCRB0111, Vero CCL-81, and Vero 76, but the copy number of chromosome 21 in Vero E6 is normal, except for the distal end. Vero E6 also showed monosomy for the X chromosome. This feature is partially observed in Vero 76, showing a mosaic copy number for the X chromosome. This indicates that the Vero 76 cell line is composed of a heterogeneous cell population, which may be distinguished by one or two copies of the X chromosome.

### Genetic Relatedness Among Sublines

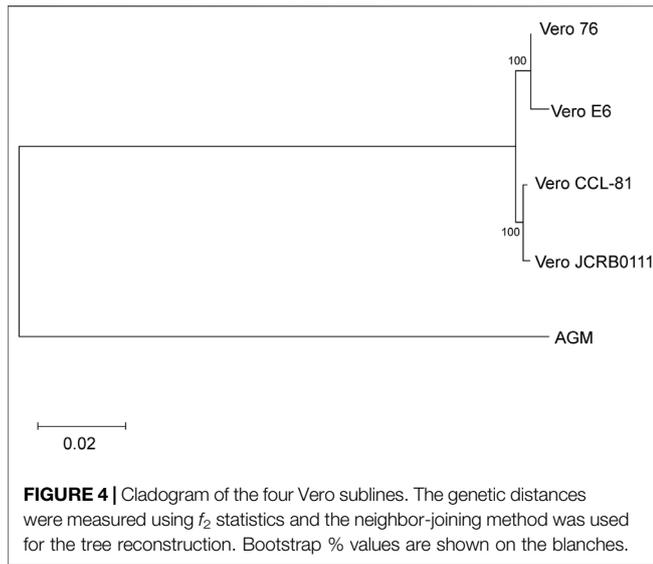
Next, we evaluated the genetic relationship among the four sublines. We used  $f_2$  statistics to measure the genetic distance between the sublines.  $F_2$  statistics are statistics that measure the difference of the frequency of variants between two cell populations (Patterson et al., 2012). The neighbor-joining tree reconstructed using  $f_2$  statistics is shown in **Figure 4**. The tree indicates the sister relationship between Vero JCRB0111 and Vero CCL-81, and between Vero 76 and Vero E6 with 100% bootstrap support. The reconstructed tree was incongruent with the inferred history of the Vero cell lineages (**Figure 1**). This may have resulted from rapid turnover of cell lineages within the sublines. The frequency of clonal cell lineages may change over time because of the difference in the survival/proliferation rate or other unknown mechanisms, and the shift would lead the differentiation of cell phenotypes (Heng and Heng, 2021). The results suggest that different lineages co-existed before the split of Vero JCRB0111 and Vero CCL-81, which increased in frequency independently in Vero JCRB0111-CCL81 and Vero 76-E6. Further studies of single-cell genomic sequencing will reveal the dynamics of the cell populations during subline divergence.

### Survey of Endogenous Retroviral Sequences

The study of Sakuma et al. (Sakuma et al., 2018) revealed many SERV sequences that are present in the genomes of Vero JCRB0111, Vero CCL-81, and Vero 76, however, the pattern of insertions was perfectly consistent among the three sublines except for one insertion, which was referred to as SVL27b. SVL27b is present in the genome of Vero 76 but absent in the genomes of Vero JCRB0111 and Vero CCL-81 (Sakuma et al., 2018). Considering that the AGM individual sequenced for draft genome harbored the insertion as a heterozygous state, Sakuma et al. (2018) concluded that the insertion was lost in Vero JCRB0111 and CCL-81 but retained in Vero 76. In this study, we investigated the pattern of SERV insertions in Vero E6, following the method of Sakuma et al. (2018) and confirmed that the integration pattern was the same as Vero 76, except for the one on the X chromosome. The SERV insertion around chrX:47012301-47012900 is heterozygous in Vero JCRB0111, Vero



**FIGURE 3 |** Copy number variation of the four Vero sublines. The dots represent the estimated copy number in 50 kb-length windows. X- and y-axes denote chromosomal coordinates and estimated copy numbers, respectively. The chromosome number is labeled under the panel. The pink boxes represent the regions of LOH. **(A)** Vero JCRB0111, **(B)** Vero CCL-81, **(C)** Vero 76, **(D)** Vero E6.



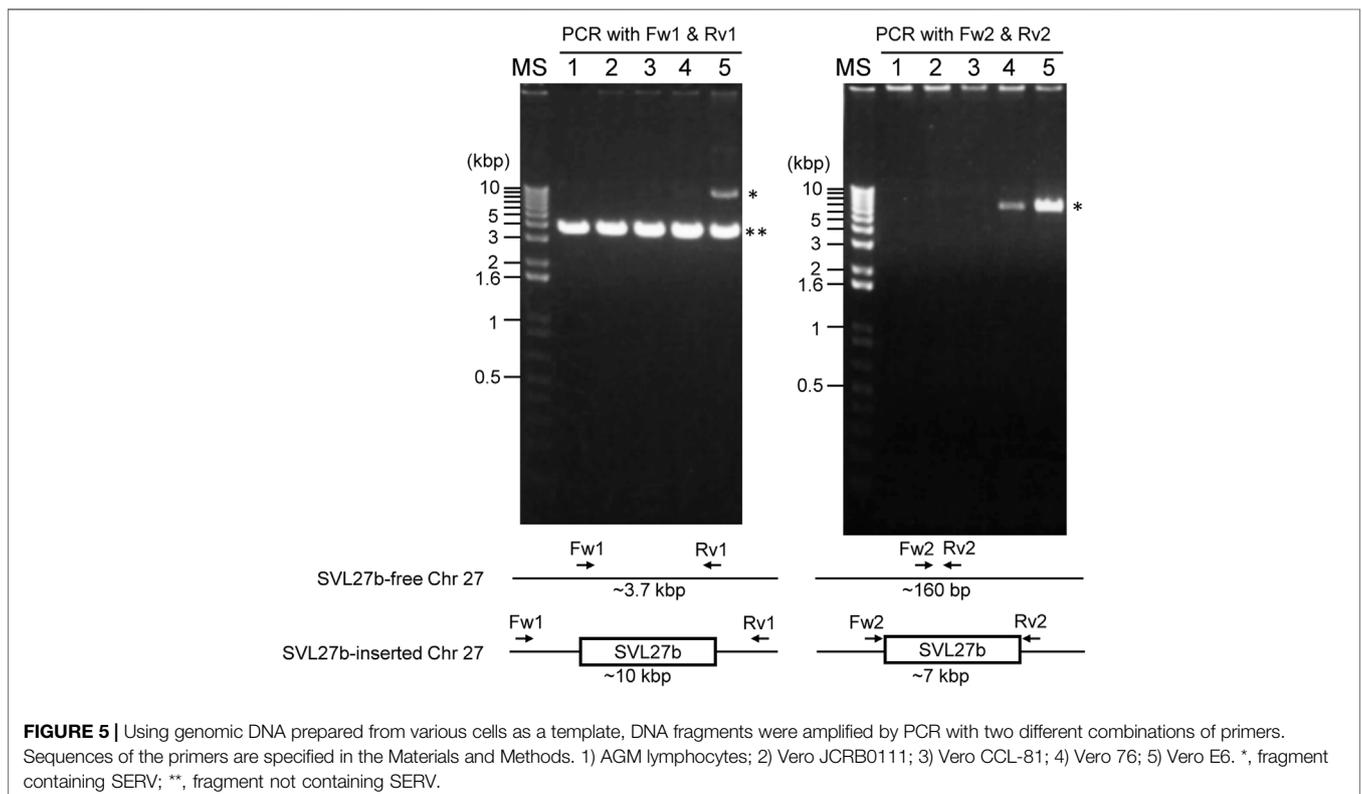
CCL-81, and Vero 76, but absent in Vero E6. The result is consistent with the CNV analysis, which showed that Vero E6 lost one of the two X chromosomes in almost every cell.

Whole-genome sequence analysis also showed that SVL27b is present in the Vero E6 genome, which is consistent with the idea that Vero E6 is a clonal derivative from Vero 76 cells (Earley and Johnson, 1988). This was verified by genomic PCR experiments as shown in **Figure 5**. When a PCR primer set matching a region ~2-

kbp away from the SVL27b inserted site, an approximate 3.7 kb DNA fragment corresponding to AGM chromosome 27, without the SERV insertion at this position, was amplified from all four Vero cell lines as well as the AGM control (**Figure 5**). This indicates that the Vero cell lines have an allele(s) that do not contain SVL27b. Notably, a DNA fragment with an ~7 kb SVL27b insertion was amplified from Vero E6 (**Figure 5**). When a PCR primer was designed proximal to the SVL27b inserted site, a DNA fragment with the SVL27b insert was amplified from Vero 76 and Vero E6 cells, but not from the others (**Figure 5**). The DNA band amplified from Vero E6 was much denser than the band from Vero 76 (**Figure 5**). It should also be pointed out that, under the latter PCR conditions, amplified DNA without the SVL27b insertion was too short to be visible using agarose gel electrophoresis. Together with the results of the previous study (Sakuma et al., 2018), these results strongly suggest that Vero 76 cells are a mixture of cell types with and without the SVL27b insertion and support that Vero E6 is derived from the Vero 76 lineage, which stably has SVL27b. Furthermore, these results indicate that SVL27b is a good genomic marker to distinguish the Vero 76-Vero E6 lineage from the Vero CCL-81 and JCRB0111 lineages (**Figure 1**).

## Identification of Subline-specific/enriched Variants

We next looked into the genetic variants specifically observed or exhibiting a high frequency in each subline. We defined a subline-specific/enriched variant as a variant with higher frequency



**TABLE 1** | Subline-specific/enriched missense variants.

	Vero JCRB0111	Vero CCL-81	Vero 76	Vero E6
Missense SNV	66	48	12	272
# of affected genes	47	35	12	188

compared with the other sublines with statistical significance (see method). In total, 22,107, 28,985, 6705, 122,198 variants were labeled as specific/enriched in Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6, respectively. Vero E6 showed the highest number of subline-specific/enriched variants. For the genes with missense SNVs, we also checked the RNA-seq expression data and retained the genes only if the mRNA with variant alleles was actually expressed in Vero E6 TMPRSS2+ cells. In total, 66, 48, 12, and 272 subline-specific/enriched missense SNVs were found on 47, 35, 12, and 188 protein-coding genes in Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6 cells, respectively (**Table 1**; **Supplementary Table S2**). In addition, we identified 4, 8, 2, and 29 genes that harbor subline-specific loss-of-function (LOF) variants resulting from SNVs and short indels in Vero JCRB0111, Vero CCL-81, Vero 76, and Vero E6 cells, respectively (**Table 2**; **Supplementary Table S3**). The analysis of large structural variants using Manta software also revealed 465, 12, 52, and 682 genes with LOH variants (**Supplementary Table S4**).

We further surveyed genes related to autophagy, apoptosis, antiviral activity, or innate immune response and harbored subline-specific/enriched missense or LOH variants in a single subline. For this purpose, we excluded the structural variants identified by Manta, because candidate variants could contain a nonnegligible number of false-positive variants and required additional experimental validations (Kawamoto et al., 2020). In total, we identified eight genes with subline-specific/enriched missense SNVs (**Table 3**). On the other hand, none of the genes with subline-specific/enriched LOH variants have functions related to the above functional categories.

## Variants in Vero E6

We focused on genetic variants specific/enriched in Vero E6, which efficiently propagate SARS-CoV-2. Missense variants in *IL1RAP* (L387F) and *BR1CC1* (D696H) were almost exclusively found in Vero E6 in a heterozygous state, whereas those in *TRIM25* (S418P) and *ATG2A* (E757Q) had a statistically high frequency in Vero E6.

*IL1RAP* is an auxiliary receptor of *IL1R1*, a receptor for interleukin (IL)-1 $\alpha$  and IL-1 $\beta$ . Virus-induced cell death releases IL-1 $\alpha$  in the cytoplasm (Malik and Kanneganti, 2018),

and the released IL-1 $\alpha$  binds to *IL1R1* on the plasma membrane. The intracellular Toll IL-1 receptor (TIR) domain of *IL1R1* binds to *IL1RAP*, leading to the activation of key transcription factors and kinases associated with the inflammatory and immune response, such as NF- $\kappa$ B, AP1, JNK, MAPK, and ERK (Mantovani et al., 2019). The leucine residue at 387 of the *IL1RAP* protein is highly conserved among mammals and prediction programs reported that the variant, leucine to phenylalanine, would be deleterious. Therefore, the L387F amino acid change in *IL1RAP* may disrupt the downstream cascade, suppressing the inflammatory and immune response and facilitating the increase in SARS-CoV-2 infection.

*RB1CC1* constitutes a part of the ULK1 complex, which is required for the initiation of autophagy (Bello-Perez et al., 2020). The ULK1 complex phosphorylates *PI3KC3*, a subunit of the phosphatidylinositol 3-kinase complex, which triggers the formation of a phagophore, and known as the sequestration membrane (Dikic and Elazar, 2018). Phagophores envelop viruses and virus-derived antigens to form autophagosomes, and the fusion of lysosomes and autophagosomes results in the formation of autolysosomes and degradation of their contents (Ahmad et al., 2018). Autophagosomes fuse with endosomes to form amphisomes, and amphisomes fuse with lysosomes to form autolysosomes (Zhao and Zhang, 2018). The endosomes may contain SARS-CoV-2 that has entered via endocytosis (Bian and Li, 2021). Previous studies have shown that ORF3a of SARS-CoV-2 inhibits two pathways that form autolysosomes, suggesting that it prevents itself from being degraded by inhibiting autophagy (Miao et al., 2021). The prediction programs suggested that the D696H variant would be deleterious and deteriorate its function. Therefore, the occurrence of non-synonymous variants in *RB1CC1* may have inhibited the initiation of autophagy and suppressed the degradation of SARS-CoV-2 by autophagy.

*TRIM25* is a ubiquitin ligase that regulates *RIG-1* (DDX58), which detects viral RNA and triggers the innate immune responses (Zeng et al., 2010). *TRIM25* causes polyubiquitination in a region of *RIG-1* called CARD, which activates *RIG-1* (Gack et al., 2007). Once activated, *RIG-1* recognizes viral RNA, it triggers the activation of NF- $\kappa$ B, IRF3, and IRF7, which induces the expression of the antiviral protein Viperin (Schneider et al., 2014). Viperin catalyzes the production of ddhCTP, which interferes with RNA polymerase (RdRp), to promote the degradation of viral proteins, and interferes with the transport of viral proteins (Rivera-Serrano et al., 2020). However, the amino acid site 418 in *TRIM25* is relatively variable among other vertebrate orthologs and

**TABLE 2** | Subline-specific LOF variants in genes related to viral infection/proliferation.

	Vero JCRB0111 sp.	VeroCCL-81 sp.	Vero 76 sp.	Vero E6 sp.
SNV	2	3	0	7 (1)
Insertion	0	0	2	4
Deletion	2	5	0	18 (2)
# of affected genes	4	8	1	29

<sup>a</sup>The numbers of homozygous variants are shown in the parentheses.

**TABLE 3** | Subline-specific/enriched missense SNVs in genes related to viral infection/proliferation.

Gene name	Vero subline	Ensembl ID	Frequency <sup>a</sup>	Amino acid change
<i>TKFC</i>	JCRB0111	ENSCSAG00000006398	23:0:0:0	G542R
<i>TGFB1</i>	JCRB0111	ENSCSAG00000003530	24:0:0:0	S61R
<i>KDR</i>	JCRB0111	ENSCSAG00000005005	21:0:0:0	V417E
<i>VAV1</i>	76	ENSCSAG00000009214	41:39:76:48	P710A
<i>IL1RAP</i>	E6	ENSCSAG00000008986	0:0:0:51	L387F
<i>TRIM25</i>	E6	ENSCSAG00000005591	40:37:49:70	S418P
<i>RB1CC1</i>	E6	ENSCSAG00000014715	1:0:0:55	D696H
<i>ATG2A</i>	E6	ENSCSAG00000007480	54:58:73:100	E757Q

<sup>a</sup>Variant frequency (%) in Vero JCRB0111:Vero CCL-81:Vero 76:Vero E6.

PROVEN software predicted that the variant of S418P would be functionally neutral.

*ATG2A* encodes a protein involved in lipid transfer between membranes (Osawa et al., 2019) and required for phagophore membrane expansion (Kotani et al., 2018). Previous studies have shown that silencing *ATG2A* results in the accumulation of unclosed autophagosomes (Velikkakath et al., 2012). The E757Q change in *ATG2A* was predicted as functionally tolerated by PROVEN and SFIT but possibly damaging by PANTHER score.

## ACE2 Gene Expression in Vero E6

We investigated the genetic variants in the *ACE2* gene of Vero E6. Sene et al. (2021) reported that the *ACE2* gene in Vero CCL-81 potentially harbors structural variants causing LOF. However, using the publicly available RNA-seq data (Zhang et al., 2021), we confirmed that *ACE2* mRNA is properly expressed in the Vero E6 TMRSS2+ cells both in the presence and absence of SARS-CoV-2. This indicates that *ACE2* is expressed in Vero cells, although the angiotensin-converting enzymatic activity of the gene product is lost as shown by Sene et al. (2021). We additionally investigated whether *ACE2* is expressed in the COS-7 cell line, which is a kidney-derived permanent cell line of AGM and was established independently of the Vero cells (Jensen et al., 1964; Gluzman, 1981). Unlike the Vero E6 cell line, a public RNA-seq data of COS-7 did not show any detectable gene expression of *ACE2*. We identified one missense SNV that showed marked differences between Vero E6 and the other sublines. The missense variants of V739I in *ACE2* were heterozygous in Vero JCRB0111, Vero CCL-81, and Vero 76, but Vero E6 harbors only the isoleucine allele. The difference results from the fact that Vero E6 exhibits monosomy X.

## CONCLUSION

In this study, we determined the whole-genome sequence of Vero E6, which has been widely used for the study of SARS-CoV-2 and performed comparative genomics on four different sublines that make up the major Vero cell lineages. Genomic resources for the Vero cell lines will benefit quality control of vaccine-producing cell substrates. In addition, finding candidate genes contributing to the different phenotypes of the cell lines will facilitate the identification of mechanisms of viral proliferation and the development of effective and safe substrates for vaccine production. The primary goal of this

study was to present a whole-genome sequence of Vero E6 as research resources and catalog a list of candidate variants that potentially affect the phenotypic differences among the Vero sublines. The validation of each effect using additional sequencing and experiments will be necessary, although it is beyond the scope of the present study. Despite of these limitations, we provide a list of genetic differences among the four sublines, as well as variants specific or enriched in particular sublines, which represent a valuable resource for quality control of cell lines and understanding the mechanisms of viral proliferation.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ddbj.nig.ac.jp/>, DRX311507.

## ETHICS STATEMENT

Ethical review and approval was not required for the animal study because we used a cell line derived from AGM. The cell line is public and the ethical review is not applicable.

## AUTHOR CONTRIBUTIONS

FK, AK, KH, TE, and NO: study design. TY and CS: performing experiments. KK and NO: data analysis. KH, FK, and NO: writing manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the MEXT KAKENHI (No. JP21H02630) and AMED-CREST (No. JP20gm0910005j0006) to KH.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.801382/full#supplementary-material>

## REFERENCES

- Ahmad, L., Mostowy, S., and Sancho-Shimizu, V. (2018). Autophagy-virus Interplay: From Cell Biology to Human Disease. *Front. Cel. Dev. Biol.* 6 155. doi:10.3389/fcell.2018.00155
- Basile, K., McPhie, K., Carter, I., Alderson, S., Rahman, H., Donovan, L., et al. (2020). Cell-based Culture Informs Infectivity and Safe De-isolation Assessments in Patients with Coronavirus Disease 2019. *Clin. Infect. Dis.* 73 (9), e2952–e2959. doi:10.1093/cid/ciaa1579
- Bello-Perez, M., Sola, I., Novoa, B., Klionsky, D. J., and Falco, A. (2020). Canonical and Noncanonical Autophagy as Potential Targets for COVID-19. *Cells* 9 (7), 1619. doi:10.3390/cells9071619
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
- Bian, J., and Li, Z. (2021). Angiotensin-converting Enzyme 2 (ACE2): SARS-CoV-2 Receptor and RAS Modulator. *Acta Pharmaceutica Sinica B* 11 (1), 1–12. doi:10.1016/j.apsb.2020.10.006
- Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., et al. (2012). Control-FREEC: A Tool for Assessing Copy Number and Allelic Content Using Next-Generation Sequencing Data. *Bioinformatics* 28 (3), 423–425. doi:10.1093/bioinformatics/btr670
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2015). Manta: Rapid Detection of Structural Variants and Indels for Germline and Cancer Sequencing Applications. *Bioinformatics* 32 (8), 1220–1222. doi:10.1093/bioinformatics/btv710
- Choi, Y., and Chan, A. P. (2015). Provean Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels. *Bioinformatics* 31 (16), 2745–2747. doi:10.1093/bioinformatics/btv195
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLOS ONE* 7 (10), e46688. doi:10.1371/journal.pone.0046688
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6 (2), 80–92. doi:10.4161/fly.19695
- Dikic, I., and Elazar, Z. (2018). Mechanism and Medical Implications of Mammalian Autophagy. *Nat. Rev. Mol. Cel Biol* 19 (6), 349–364. doi:10.1038/s41580-018-0003-4
- Earley, E., and Johnson, K. (1988). “The Lineage of the Vero, Vero 76 and its Clone C1008 in the United States,” in *VERO Cell Origin, Properties Biomedical Application*. Editors B. Simizu and T. Terasima (Chiba: Chiba University), 26–29.
- Gack, M. U., Shin, Y. C., Joo, C.-H., Urano, T., Liang, C., Sun, L., et al. (2007). TRIM25 RING-finger E3 Ubiquitin Ligase Is Essential for RIG-I-Mediated Antiviral Activity. *Nature* 446 (7138), 916–920. doi:10.1038/nature05732
- Gluzman, Y. (1981). SV40-transformed Simian Cells Support the Replication of Early SV40 Mutants. *Cell* 23 (1), 175–182. doi:10.1016/0092-8674(81)90282-8
- Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg. Infect. Dis.* 26 (6), 1266–1273. doi:10.3201/eid2606.200516
- Heng, J., and Heng, H. H. (2021). Karyotype Coding: The Creation and Maintenance of System Information for Complexity and Biodiversity. *Biosystems* 208, 104476. doi:10.1016/j.biosystems.2021.104476
- Jensen, F. C., Girardi, A. J., Gilden, R. V., and Koprowski, H. (1964). Infection of Human and Simian Tissue Cultures with Rous Sarcoma Virus. *Proc. Natl. Acad. Sci.* 52 (1), 53–59. doi:10.1073/pnas.52.1.53
- Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: A Stateful Web Service to Facilitate Gene/protein List Analysis. *Bioinformatics* 28 (13), 1805–1806. doi:10.1093/bioinformatics/bts251
- Kawamoto, M., Yamaji, T., Saito, K., Shirasago, Y., Satomura, K., Endo, T., et al. (2020). Identification of Characteristic Genomic Markers in Human Hepatoma HuH-7 and Huh7.5.1-8 Cell Lines. *Front. Genet.* 11 (1256), 546106. doi:10.3389/fgene.2020.546106
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi:10.1038/s41587-019-0201-4
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing. *Genome Res.* 22 (3), 568–576. doi:10.1101/gr.129684.111
- Kotani, T., Kirisako, H., Koizumi, M., Ohsumi, Y., and Nakatogawa, H. (2018). The Atg2-Atg18 Complex Tethers Pre-autophagosomal Membranes to the Endoplasmic Reticulum for Autophagosome Formation. *Proc. Natl. Acad. Sci. USA* 115 (41), 10363–10368. doi:10.1073/pnas.1806727115
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi:10.1093/molbev/msy096
- Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Malik, A., and Kanneganti, T.-D. (2018). Function and Regulation of IL-1 $\alpha$  in Inflammatory Diseases and Cancer. *Immunol. Rev.* 281 (1), 124–137. doi:10.1111/imr.12615
- Mantovani, A., Dinarello, C. A., Molgora, M., and Garlanda, C. (2019). Interleukin-1 and Related Cytokines in the Regulation of Inflammation and Immunity. *Immunology* 50 (4), 778–795. doi:10.1016/j.immuni.2019.03.012
- Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., et al. (2020). Enhanced Isolation of SARS-CoV-2 by TMPRSS2-Expressing Cells. *Proc. Natl. Acad. Sci. USA* 117 (13), 7001–7003. doi:10.1073/pnas.2002589117
- Miao, G., Zhao, H., Li, Y., Ji, M., Chen, Y., Shi, Y., et al. (2021). ORF3a of the COVID-19 Virus SARS-CoV-2 Blocks HOPS Complex-Mediated Assembly of the SNARE Complex Required for Autolysosome Formation. *Dev. Cel.* 56 (4), 427–442. e425. doi:10.1016/j.devcel.2020.12.010
- Mizusawa, H., Simizu, B., and Terasima, T. (1988). “Cell Line Vero Deposited to Japanese Cancer Research Resources Bank,” in *VERO Cell Origin, Properties Biomedical Application*. Editors B. Simizu and T. Terasima (Chiba: Chiba University), 24–25.
- Ogando, N. S., Dalebout, T. J., Zevenhoven-Dobbe, J. C., Limpens, R. W. A. L., van der Meer, Y., Caly, L., et al. (2020). SARS-coronavirus-2 Replication in Vero E6 Cells: Replication Kinetics, Rapid Adaptation and Cytopathology. *J. Gen. Virol.* 101 (9), 925–940. doi:10.1099/jgv.0.001453
- Osada, N., Kohara, A., Yamaji, T., Hirayama, N., Kasai, F., Sekizuka, T., et al. (2014). The Genome Landscape of the African green Monkey Kidney-Derived Vero Cell Line. *DNA Res.* 21 (6), 673–683. doi:10.1093/dnares/dsu029
- Osawa, T., Kotani, T., Kawaoka, T., Hirata, E., Suzuki, K., Nakatogawa, H., et al. (2019). Atg2 Mediates Direct Lipid Transfer Between Membranes for Autophagosome Formation. *Nat. Struct. Mol. Biol.* 26 (4), 281–288. doi:10.1038/s41594-019-0203-4
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient Admixture in Human History. *Genetics* 192 (3), 1065–1093. doi:10.1534/genetics.112.145037
- Rivera-Serrano, E. E., Gizzi, A. S., Arnold, J. J., Grove, T. L., Almo, S. C., and Cameron, C. E. (2020). Viperin Reveals its True Function. *Annu. Rev. Virol.* 7 (1), 421–446. doi:10.1146/annurev-virology-011720-095930
- Saito, K., Fukasawa, M., Shirasago, Y., Suzuki, R., Osada, N., Yamaji, T., et al. (2020). Comparative Characterization of Flavivirus Production in Two Cell Lines: Human Hepatoma-Derived Huh7.5.1-8 and African green Monkey Kidney-Derived Vero. *PLOS ONE* 15 (4), e0232274. doi:10.1371/journal.pone.0232274
- Sakuma, C., Sekizuka, T., Kuroda, M., Kasai, F., Saito, K., Ikeda, M., et al. (2018). Novel Endogenous Simian Retroviral Integrations in Vero Cells: Implications for Quality Control of a Human Vaccine Cell Substrate. *Sci. Rep.* 8 (1), 644. doi:10.1038/s41598-017-18934-2
- Schneider, W. M., Chevillotte, M. D., and Rice, C. M. (2014). Interferon-stimulated Genes: A Complex Web of Host Defenses. *Annu. Rev. Immunol.* 32 (1), 513–545. doi:10.1146/annurev-immunol-032713-120231
- Sène, M.-A., Kiesslich, S., Djambazian, H., Ragoussis, J., Xia, Y., and Kamen, A. A. (2021). Haplotype-resolved De Novo Assembly of the Vero Cell Line Genome. *npj Vaccin.* 6 (1), 106. doi:10.1038/s41541-021-00358-9

- Tang, H., and Thomas, P. D. (2016). Panther-psep: Predicting Disease-Causing Genetic Variants Using Position-specific Evolutionary Preservation. *Bioinformatics* 32 (14), 2230–2232. doi:10.1093/bioinformatics/btw222
- Terasima, T., Yasukawa, M., and Simizu, B. (1988). “History of Vero Cells in Japan,” in *VERO Cell Origin, Properties Biomedical Application*. Editors B. Simizu and T. Terasima (Chiba: Chiba University), 22–23.
- Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics* 5, 99–114. doi:10.2307/3001913
- Velikkakath, A. K. G., Nishimura, T., Oita, E., Ishihara, N., and Mizushima, N. (2012). Mammalian Atg2 Proteins Are Essential for Autophagosome Formation and Important for Regulation of Size and Distribution of Lipid Droplets. *MBoC* 23 (5), 896–909. doi:10.1091/mbc.e11-09-0785
- Woolsey, C., Borisevich, V., Prasad, A. N., Agans, K. N., Deer, D. J., Dobias, N. S., et al. (2021). Establishment of an African green Monkey Model for COVID-19 and protection against Re-infection. *Nat. Immunol.* 22 (1), 86–98. doi:10.1038/s41590-020-00835-8
- Zeng, W., Sun, L., Jiang, X., Chen, X., Hou, F., Adhikari, A., et al. (2010). Reconstitution of the RIG-I Pathway Reveals a Signaling Role of Unanchored Polyubiquitin Chains in Innate Immunity. *Cell* 141 (2), 315–330. doi:10.1016/j.cell.2010.03.029
- Zhang, Y., Guo, R., Kim, S. H., Shah, H., Zhang, S., Liang, J. H., et al. (2021). SARS-CoV-2 Hijacks Folate and One-Carbon Metabolism for Viral Replication. *Nat. Commun.* 12 (1), 1676. doi:10.1038/s41467-021-21903-z
- Zhao, Y. G., and Zhang, H. (2018). Autophagosome Maturation: An Epic Journey from the ER to Lysosomes. *J. Cel. Biol.* 218 (3), 757–770. doi:10.1083/jcb.201810099
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Konishi, Yamaji, Sakuma, Kasai, Endo, Kohara, Hanada and Osada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.