Nanopore sequencing and *de novo* assembly of a misidentified Camelpox vaccine reveals putative epigenetic modifications and alternate protein signal peptides

Zack Saud (zack.saud@swansea.ac.uk)
Swansea University

Matthew D. Hitchings
Swansea University

Tariq M. Butt
Swansea University

Research Article

**Keywords:** Camelpox, Vaccinia, Epigenomics, Epigenetics, WGS

**Posted Date:** October 8th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-89149/v1

**License:** © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

DNA viruses can exploit host cellular epigenetic processes to their advantage; however, the epigenome status of most DNA viruses remains undetermined. Third generation sequencing technologies allow for the identification of modified nucleotides from sequencing experiments without specialized sample preparation, permitting the detection of non-canonical epigenetic modifications that may distinguish viral nucleic acid from that of their host, thus identifying attractive targets for advanced therapeutics and diagnostics. We present a novel nanopore de novo assembly pipeline used to assemble a misidentified Camelpox vaccine. Two confirmed deletions of this vaccine strain in comparison to the closely related Vaccinia virus strain modified vaccinia Ankara make it one of the smallest non-vector derived orthopoxivirus genomes to be reported. Annotation of the assembly revealed a previously unreported signal peptide at the start of protein A38 and several signal peptides that were found to differ from those previously described. Putative epigenetic modifications around various motifs have been identified and the assembly confirmed previous work showing the vaccine genome to most closely resemble that of Vaccinia virus strain Modified Vaccinia Ankara. The pipeline may be used for other DNA viruses, increasing the understanding of DNA virus evolution, virulence, host preference, and epigenomics.

Introduction

DNA viruses include those which have DNA genomes and replicate using DNA-dependent DNA polymerase. They are grouped into two classes, comprising single stranded DNA viruses and double stranded DNA viruses. The latter group contains the infamous Variola Virus (VARV), the causative agent of smallpox, which belongs to the family Poxviridae, subfamily Chordopoxvirinae and genus Orthopoxvirus. There are currently 10 accepted species within the genus, the other notable members including; Vaccinia virus (VACV) – the prototype Orthopoxvirus used as a vaccine to eradicate human smallpox and which has no known natural host [1], Cowpox virus (CPXV) - administered successfully by Edward Jenner as the first documented successful vaccine [2], Monkeypox virus (MPXV) – a zoonotic virus endemic to the African subcontinent [3], and Camelpox (CMLV) – the most genetically similar extant species to VARV [4].

Poxviruses have linear, double-stranded DNA genomes that vary from 130 to 230 kbp [5]. The telomere ends of the genome form covalently closed hairpin structures at the termini [6]. The hairpin is at the end of a long, inverted terminal repetition (ITR) containing sets of short, tandemly repeated sequences [5]. For orthopoxviruses, the size of the ITRs range from approximately 200 to 500 base pairs for variola viruses, to almost 12,000 base pairs for several vaccinia virus strains [7]. Large ITR regions can pose problems for first generation Sanger sequencing [8] and second-generation Illumina sequencing [9], which are capable of producing sequence read lengths of up to around 1000 bp and 300 bp (or around 500 bp linked pair-end) respectively. Such tracts of repetitive sequences in a genome can be resolved by third-generation long read sequencing technologies [10-12], which are capable of producing read lengths in excess of 100,000 bp.

The central portions of most poxvirus genomes are highly conserved, and contain essential genes involved in key functions such as transcription, DNA replication and virion assembly [13]. In contrast, genes that cluster at the ends of the genome are usually species or host specific, and encode virulence factors that modulate the host immune system [13,14]. Various proteins encoded by the genome have been shown to interact with DNA or precursor nucleotides [5]. The K7 protein has been shown to promote histone methylation associated with heterochromatin formation [15]. Furthermore, vaccinia virus (VACV) C4 [16], C6 [17], C16 [18], B14 [19], E3 [20], F16 [21], and N2 [22] gene products can be detected in the host nucleus, thus implicating them in some form of transcriptional regulation. To our knowledge, no research has been aimed towards assessing whether these proteins epigenetically modify the viral DNA. Furthermore, despite what is known of the capability of DNA viruses to exploit host cellular epigenetic processes to their advantage during infection [23,24], the epigenome status of most DNA viruses remains unknown.

Third generation sequencing technologies have advanced epigenomic research by providing platforms that allow for the identification of modified nucleotides from sequencing experiments without the need for specialized secondary sample preparation protocols [25-27]. Such a direct approach for interrogating an epigenome is particularly beneficial for viral epigenetic research, as samples often contain high amounts of contaminating host DNA, which can complicate specialized DNA methylation probing techniques such as bisulfite sequencing [28] and antibody based approached [29]. Furthermore, in the case of Nanopore sequencing, motifs with non-canonical epigenetic modifications can be identified by distinguishing a deviation of the raw signal from that of a standard model at a given nucleotide sequence [26,30]. Such non-canonical epigenetic modifications would distinguish viral DNA from that of host DNA, making them attractive targets for advanced therapeutics and diagnostics [31].

In this study, we use nanopore sequencing to assemble the genome of a live attenuated CMLV strain, Ducapox, that was stated to comprise a CMLV isolate from the United Arab Emirates (CaPV298-2) [32]. The vaccine has since been found to contain two gene regions that more closely resembled that of VACV strain Modified Vaccinia Ankara (VACV-MVA) [33]. A separate study of the strain using second generation WGS found the vaccine genome matches that of VACV-MVA, with the exception of two genomic deletions (5195 and 890 bp in size), however, the authors questioned the authenticity of these genomic deletions due to both the reference-based assembly approach adopted, and the low sequencing coverage of the genome [34]. We present a sequencing and annotation pipeline for long read de novo assembly of Poxvirus genomes and identify putative epigenetic modifications within the genome. Using the latest version of signal peptide identification software, we identify a protein with a
previously undescribed signal peptide, and present several signal peptides that were found to differ from previously described sequences. The pipeline may be used for other DNA viruses, increasing the understanding of DNA virus epigenomics.

**Results**

**Sequencing statistics and de novo assembly**

Table 1 lists the read output metrics of the Nanopore MinION. Base calling produced a total of 405,925 sequences, of which 16,059 (3.95%) remained after size filtering and removal of contamination. Most of the background contamination was found to be simian DNA, consistent with the virus having been propagated in Vero cells. The assembler initially produced a contig that was 195,695 bp in length. After all polishing steps, the final genome was 159,696 bp. Figure 2 shows read map coverage distributions across both the initial assembly and the final polished assembly. Read coverage was found to be more uniformly distributed in the final assembly in comparison to the initial assembly, the latter of which was found to have uneven read coverage distributions at the contig ends. This is indicative of the final polished assembly containing terminal repeat sequence lengths that more closely match that of ground truth.

**Whole genome sequence comparisons**

A blast search of the final polished assembly revealed the genome to most closely match that of Vaccinia virus strain Acambis 3000 MVA (Genbank accession: AY603355.1), with a percentage identity score of 99.99%. Figure 3 shows a dotplot comparison of the Ducapox long read assembly vs the closest matching viral genome, that of VACV Acambis 3000 MVA. Genomic deletions of 5449 bp and 916 bp in size were seen in the Ducapox genome, corresponding to VACV Acambis 3000 MVA genome positions 3735 – 9183, and 23,219 – 24,134, respectively. These deletions were confirmed by visualizing the mapping of reads to the genome assembly, and confirming that unbroken reads traversed the deletion sites (supplementary information 1a and 1b). The VACV Acambis 3000 MVA was also found to be 227 bp and 435 bp longer at its ends, with respect to the Ducapox genome. Supplementary information 1c features a multiple sequence alignment between the Ducapox long read genome assembly, the Ducapox short read genome assembly, and the VACV Acambis 3000 MVA genome.

**Genome annotations and functional analyses**

Figure 4 illustrates the annotated Ducapox gene map. The genome was found to contain a total of 191 protein coding genes. A total of 200 genes were initially predicted by Prodigal, however, 9 of these predicted genes were found to contain no functional domain, and had no significant percentage identity to any protein in the Swissprot database, hence were removed from subsequent analyses. Table 2 lists proteins containing signal peptides, which included a total of 13 out of the 191 proteins predicted. A comparison of the proteins predicted by SignalP v5.0 (the latest version) and the signal peptides listed in the Uniprot database revealed that SignalP v5.0 predicted one previously unreported signal peptide in the protein A38L. Two proteins (A39R and HA) were found to have signal peptides predicted by SignalP v5.0 that matched those in the Uniprot database. The remaining 10 proteins contained signal peptides predicted by SignalP v5.0 that differed from those in the Uniprot database. Supplementary information 2 contains the SignalP v5.0 output summary. Supplementary information 3 contains a list of mature proteins with the SignalP 5.0 predicted signal peptides removed. StructRNAfinder identified a single structural RNA- the Pox_AX_element (RF00385), involved in directing the efficient production and orientation-dependent formation of late RNAs.

**Putative epigenetically modified sites**

A total of four putative motifs were identified in the Ducapox genome. Figure 5 shows the statistical plot for the AGAAGRC, as well as the sequence logo. Figure 5 shows the statistical plot and sequence logo of the AGAAGRC motif. The statistical plot is based on 17 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model were observed around the central AAG nucleotides. A Tomtom search of the motif detected no similar known motifs. Figure 6 shows the statistical plot and sequence logo of the AARRRGATKH motif. The statistical plot is based on 42 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model were observed around the central GA nucleotides. A Tomtom search of the motif showed the reverse-complement to most closely match MA0467.1 (Crx binding motif; Mus musculus) in the JASPAR database. A DELTA-BLAST of the CRX_Mouse (>sp|O54751|CRX_MOUSE) amino acid sequence using Vaccinia virus as the organism revealed low alignment scores (<40) of several Ducapox proteins to the Homeobox (DNA-binding) domain within the CRX protein that included; the RPO147 gene product- DNA-directed RNA polymerase 147 kDa polypeptide, the G9R gene product- myristoylated protein G9, the E8R gene product- protein E8, and part of the L3L gene product. A DELTA-BLAST of the KLF4 (>sp|O43474|KLF4_HUMAN) amino acid sequence using Vaccinia virus as the organism revealed a low alignment score (<40) of a single Ducapox protein to the putative nucleic acid binding site within the KLF4 protein; gene product A3L- the major coat protein 4b.
Figure 8 shows the statistical plot and sequence logo of the WWAATGWC motif. The statistical plot is based on 77 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model were observed around the central TGT nucleotides. A Tomtom search of the motif showed the reverse-complement to most closely match MA1112.1 (NR4A1; Homo sapiens) in the JASPAR database. A DELTA-BLAST of the NR4A1 (>sp|P22736|NR4A1_HUMAN) amino acid sequence using Vaccinia virus as the organism revealed low alignment scores (<40) of a single Ducapox protein to the putative nucleic acid binding site within the NR4A1 protein; gene product F3L- the Kelch repeat protein F3.

Discussion

Except for two confirmed genomic deletions, the whole genome sequence of this vaccine was shown to closely resemble that of VACV-MVA, supporting our earlier study in which we reported that two gene regions of this vaccine resembled the aforementioned strain [33]. Our findings also corroborate with a previous study that used short read Illumina sequencing, and a reference guided assembly to generate a partial Ducapox genome, wherein the authors noted the putative deletions, but could confirm the validity of the deletions due to the both the assembly pipeline and sequencing technology used [34]. At 159,696 bp in length, the vaccine genome, to our knowledge, is the smallest amongst the non-vector derived orthopoxviruses. We postulate that the deletions may have been a result of passage of a misidentified VACV-MVA strain, as it is known that poxvirus genomes tend to decrease in size with serial passage [35]. It has been demonstrated that VACV has a defined origin of replication, which supports a model for poxvirus genome replication that involves leading and lagging strand synthesis [36]. Studies on poxvirus DNA replication described putative Okazaki fragments of about 1,000 nt in length (suspiciously similar in size to the 916 bp deletion of the Ducapox sequence) and RNA primers on the 5'-ends of newly made chains of VACV DNA [37, 38].

We identified a previously unreported signal peptide in protein A38L. The A38L gene product is a 33kDa integral membrane glycoprotein [39]. Overexpression of the protein has been shown to promote Ca^{2+} influx into infected cells [40]. The latest version of SignalIP predicted alternate peptide signals for 10 other proteins. These include; the gene product of C8L- the function of which remains unknown, the gene product of B19R- a type 1 interferon decoy [41], the gene product of E10R- associated with membranes of intracellular mature virions and plays a role in morphogenesis [42], the gene product of B8R- another interferon decoy [41], the gene product of B7R- which is involved with virulence [43], the gene product of B16R- an IL-1β binding protein [44], the gene product of SPI-3- a cell fusion inhibitor protein [45], the gene product of PS/HR- which plays a role in the dissolution of the outermost membrane of extracellular enveloped virions to allow virion entry into host cells and also participates in wrapping mature virions to form enveloped virions [46], and finally the gene product of A43R- which enhances intradermal lesion formation [47]. Signal peptides play a range of different roles within the cells that include marking proteins for secretion, intracellular translocation, and keeping cells inactive as precursor proteins until signal peptide cleavage [48]. Further research is needed to determine whether biochemical analyses of these new mature proteins yields any further insight into protein function.

We have presented, for the first time to our knowledge, evidence of epigenetic modifications in a Poxvirus genome. We have identified motifs within which there are bases have signals that are consistently different to a canonical model, and we've identified similar motifs, the proteins of which contain DNA binding domains with some homology to various Vaccinia virus proteins. Although the Nanopore is a valuable tool for identifying putative epigenetic sites in a genome, the device does not allow for the identification of either the particular base that is modified (sometimes it is neighboring bases to the epigenetically modified base that show signal divergence from the canonical model), nor does it allow for the identification of the modifying chemical group. Thus, further analyses are required to confirm the results. Modifications that distinguish viral DNA from that of the host may be targets for advanced therapeutics. Should these epigenetic modifications be confirmed and chemically characterized, another important question would concern whether the modifications were the result of a viral protein, or the result of a host protein.

Given the relative cheapness of Nanopore sequencing, future research could investigate the evolutionary trajectory of orthopoxviruses with continued passage. Experiments such as determining whether different evolutionary trajectories occur when a seed stock of a virus is passaged in differing permissive cell lines would be of great interest. Furthermore, the Nanopore would allow for the assessment of differing epigenome modifications with continued passage. Such studies would assist in providing further evidence towards efforts to better understand the origins of Vaccinia virus [49]. Additionally, long read sequencing transcriptomics techniques have recently shed light on the high variation in transcript lengths at certain Vaccinia genome loci, termed chaotic regions [50,51]. Long read sequencing coupled with these transcriptomics techniques could provide greater insight into the loss of Poxvirus virulence with passage.

To conclude, we have developed a novel assembly pipeline for long read sequencing of Poxvirus genomes, that corrects the lengths of terminal ends. The two confirmed deletions of this vaccine strain in comparison to VACV-MVA make it one of the smallest non-vector derived orthopoxvirus genomes to be reported. We have used the latest software for signal peptide prediction to discover a novel signal peptide in a VACV protein that has not been previously reported, as well as discovering 10 alternate signal peptides in comparisons to those previously reported. We have presented the first evidence of epigenetic modifications in a Poxvirus, and detected motifs, identified similar motifs with known binding proteins, and shown that the binding domains of these known proteins share some homology to those of various Ducapox proteins. The methods we have detailed may be used for other viral genomes, thus aiding the understanding of the molecular mechanisms underpinning viral virulence, evolution and host preferences.
Methods

Source and composition of vaccine

A commercial live attenuated 'Ducapox' vaccine was sourced from Al Bashayer Veterinary Supplies (Dubai, United Arab Emirates), manufactured by Design Biologix (Pretoria, South Africa) and commercialized by Highveld Biological Ltd (Johannesburg, South Africa). The CMLV strain CaPV298-2, the parent strain of this vaccine, was originally isolated in the United Arab Emirates and attenuated through serial passage in Vero cell culture [32]. Manufacture and expiry dates were 07–2018 and June 2019, respectively and the batch number was DPV0818.

DNA extraction

DNA was extracted using the QIAamp DNA Mini kit (Catalog # 51304, Qiagen, Hilden, Germany), following the DNA purification from tissues protocol, adding 180 µL of Buffer ATL to 25 mg of lyophilized vaccine and following the manufacturer's guidelines with the addition of adding 5 µg of Carrier RNA Poly A (Catalog # 1017647, Qiagen, Hilden, Germany) to the 200 µL of Buffer AL solution. The DNA preparation was analyzed for purity on a nanodrop spectrophotometer (ThermoScientific, Rochester, USA), and the concentration was determined using a Qubit dsDNA assay kit (ThermoScientific, Rochester, USA) and a Qubit 4 fluorometer (ThermoScientific, Rochester, USA).

Preparation of nanopore library and sequencing

400 ng of genomic DNA was used for Nanopore library preparation using a Rapid Sequencing Kit (SQK-RAD004, Oxford Nanopore Technologies) and barcode 18 of the Native Barcoding Expansion kit (EXP-NBD114, Oxford Nanopore Technologies). Multiplexed sequencing was performed on a MinION device (Oxford Nanopore Technologies), equipped with a R9.4.1 MinION flow cell. Base calling was performed offline with ONT’s Guppy software pipeline version 4.0.11, enabling the --pt_scaling flag, setting --trim_strategy to DNA, loading the dna_r9.4.1_450bps_hac configuration files, and setting --barcode_kits EXP-NBD114.

Long read pre-processing, assembly, and polishing

Long read adapter trimming was performed with Porechop version 0.2.4 (www.github.com/rrwick/Porechop), setting both the --adapter_threshold and --barcode_threshold to 98. The trimmed long reads were filtered to remove reads under 3000 bases in length using NanoFilt version 2.6.0 [52]. The adapter trimmed, filtered long reads initially assembled using Flye version 2.8 [53] using the --nano-raw, --meta, --trestle and --keep-haplotype flags. A fasta file of non-viral assembled contigs (identified using a blast search) was made from the assembly output using Bandage version 0.8.1 [54]. The adapter trimmed, filtered long reads were mapped to the non-viral assembled contigs using minimap2 version 2.17-r941 [55], and the unmapped reads were extracted from the alignment file and converted to FASTQ using samtools [56], this generating a read set exclusively containing viral DNA. The virus specific reads were assembled using Flye version 2.8, enabling the --nano-raw, setting the minimum overlap to 5000 using the -m 5000 flag, and conducting 3 polishing iterations by setting the -i 3 flag. The assembly was polished, correcting the ITR regions, using the --only-polish flag of the tandemquast tool of the TandemTools package [57]. Long reads were mapped to the assembly using minimap2 version 2.17-r941, and the resulting alignment file was used to polish the assembly with Racon version v1.4.13 [58] using the following parameters: -m 8 -x -6 -g -8 -w 500 -no-trimming. A total of 3 rounds of mapping and polishing with Racon were done on the assembly, after which no changes were observed. The corrected consensus was further polished with the same long read set using Medaka version 0.11.5 (https://github.com/nanoporetech/medaka), setting the -m r941_min_high_g360 flag. Figure 1 shows a graphical representation of the full assembly pipeline.

Assessment of assemblies and whole genome comparisons

The contamination-free, adapter trimmed, filtered long reads were mapped to both the initial Flye assembly, and the final polished assembly in order to manually assess for the absence of read mapping breaks by plotting read mapping coverage of genome assemblies using pyGenomeTracks version 3.5 [59]. Genome comparisons were performed using the nucmer tool of Mummer 3 [60]. The final polished assembly was compared against the shot-read Ducapox assembly (Genbank accession: MT648498.1) and Vaccinia virus strain Acambis 3000 MVA (Genbank accession: AY603355.1), the closest matching genome to the long-read assembly as determined by an online BLAST search.

Genome annotation

The polished assembly was annotated using Prodigal v2.6.3 [61]. The annotation gff3 file was loaded into the GenSAS suite version 6.0 [62], after which functional analyses was conducted in the suite using InterProScan version 5.25-68.0 [63] and the ab initio predicted proteins were identified using blastp [64] by conducting a protein vs protein search against the SwissProt protein data set to determine best matches. Protein sequences were analyzed for signal peptides using the SignalP v5.0 [65]. Non-coding RNAs were detected using StructRNAfinder [66].

Putative epigenetic modification detection
Nanopore raw signals were analyzed for divergence from a standard model, signifying putative epigenetic modification sites, using the Tombo suite [26]. After running Tombo's resquiggle function using the final polished genome, the detect_modifications function was run using the de_novo model. The results of the stats file was converted to a FASTA file using the text_output function of Tombo, setting –num-regions 1000 and –num-bases 15. The central 7 nucleotides of each entry of the fasta file was plotted using the motif_with_stats function in Tombo, using the maximum –num-statistics number that would produce a plot for each fasta entry (determined empirically) for all entries with scores >0.7 for Frac. Alternate in the fasta file. The motif_with_stats plots were assessed manually, and the motifs from plots containing increases in the fraction of modified bases (-log10(P-value) exclusively around the central motif only were kept, and these were used to create a separate fasta file containing all motifs for each of the four modified bases that were manually detected from the plots. Meme v5.1.1 [67] was used on each individual fasta file using the -dna and -mod zoops ags to determine motifs. Motifs were compared to known motifs using Tomtom v5.1.1 [68], and the amino acid sequence of the protein that bound to the most similar motif was used as input for a BlastP online search against the UniProtKB/Swiss-Prot database, using Vaccinia virus (taxid:10245) as the organism, and using the DELTA-BLAST algorithm so as to determine putative VACV proteins with similar domains that may bind the same motifs.

**Declarations**

**Acknowledgements**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank staff at the University of Sharjah for facilitating the use of their facilities for DNA extraction.

**Author information**

Zack Saud, Matthew D. Hitchings and Tariq M. Butt contributed equally.

**Contributions**

Z.S. performed DNA extraction, bioinformatics analyses, and wrote the manuscript. M.D.H. performed Nanopore sequencing. T.M.B provided oversight, reviewed the manuscript and provided laboratory support.

**Correspondence to Zack Saud**

**Data availability**

All data generated in this study has been deposited at the NCBI under Bioproject PRJNA663037. Nanopore sequencing read data can be accessed at the NCBI SRA using the accession number SRR12667950. Sample information can be accessed at the NCBI BioSample repository using the accession number SAMN16115327. The genome assembly can be accessed using the accession number MT946551. Gene and protein names, and functional annotations (GO terms, InterPro, PFAM) are included in GenBank entries. Bioinformatics tool output files have been deposited in the following GitHub repository- https://github.com/zacksaud/Ducapox-Assembly-Project, as well as in the supplementary information.

**Ethics declarations**

**Competing Interests**

The authors declare no competing interests.

**References**

1. Fenner, F., Henderson, D.A., Arita, I., Jezek, Z. & Ladnyi, I.D. Smallpox and its eradication. Geneva: World Health Organization; 1988. [March 14, 2003]. p. 1460. Reference out-of-print. See the World Health Organization, Communicable Disease Surveillance and Response Web site. Available at: www.who.int/emc/diseases/smallpox/smallpoxeradication.html.

2. Jenner, E. An Inquiry Into the Causes and Effects of the Variole Vaccinæ, a Disease Discovered in Some of the Western Counties of England, Particularly Gloucestershire and Known by the Name of the cow-pox. London: Sampson Low, 1798.

3. Sklenovská, N. & Van Ranst, M. Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. *Front Public Health* 6, 241. https://doi.org/10.3389/fpubh.2018.00241 (2018)

4. Gubser, C. & Smith, G.L. The sequence of camelpox virus shows it is most closely related to variola virus, the cause of smallpox. *J Gen Virol* 83, 855–872. https://doi.org/10.1099/0022-1317-83-4-855 (2002)

5. Moss, B. Poxvirus DNA replication. *Cold Spring Harb Perspect Biol* 5(9), a010199 https://doi.org/10.1101/cshperspect.a010199 (2013)

6. Winters, E., Baroudy, B.M. & Moss, B. Molecular cloning of the terminal hairpin of vaccinia virus DNA as an imperfect palindrome in an *Escherichia coli* plasmid. *Gene* 37, 221–228 https://doi.org/10.1016/0378-1119(85)90276-8 (1985)
7. Hendrickson, R.C., Wang, C., Hatcher, E.L. & Lefkowitz, E.J. Orthopoxvirus genome evolution: the role of gene loss. *Viruses* 2(9), 1933-1967 https://doi.org/10.3390/v2091933 (2010)

8. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12), 5463-5467 https://doi.org/10.1073/pnas.74.12.5463 (1977)

9. Bennett, S. Solaexa ltd. *Pharmacogenomics* 5(4), 433-438. https://doi.org/10.1517/14622416.5.4.433 (2004)

10. Kasiyanowicz, J.J., Brandin, E., Branton, D. & Deamer, D.W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA* 93 3770-3773 https://doi.org/10.1073/pnas.93.24.3770 (1996)

11. Jain, M., Olsen, H.E., Paten, B., Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17(1) 239 https://doi.org/10.1186/s13059-016-1103-0 (2016)

12. Eid, J., Fehr, A., Gray, J., et al. Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910) 133-138. https://doi.org/10.1126/science.1162986 (2009)

13. Gubser, C., Hué, S., Kellam, P. & Smith, G.L. Poxvirus genomes: a phylogenetic analysis. *J Gen Virol* 85(1), 105-117. https://doi.org/10.1099/vir.0.19565-0 (2004)

14. Moss, B. Poxviridae: the viruses and their replication. In Fields Virology, 4th edition, pp. 2849–2883. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins (2001)

15. Teferi, W.M., Desauniers, M.A., Noyce, R.S., Shenouda, M., Umer, B. & Evans, D.H. The vaccinia virus K7 protein promotes histone methylation associated with heterochromatin formation. *PLoS One* 12(3), e0173056 https://doi.org/10.1371/journal.pone.0173056 (2017)

16. Ember, S.W., Ren, H., Ferguson, B.J. & Smith, G.L. Vaccinia virus protein C4 inhibits NF-κB activation and promotes virus virulence. *J Gen Virol* 93(10), 2098–2108 https://doi.org/10.1099/vir.0.045070-0 (2012)

17. Unterholzner, L., Sumner, R.P., Baran, M., Ren, H., Mansur, D.S., Bourke, N.M. et al. Vaccinia virus protein C6 is a virulence factor that binds TBK-1 adaptor proteins and inhibits activation of IRF3 and IRF7. *PLoS Pathog* 7(9), e1002247 https://doi.org/10.1371/journal.ppat.1002247 (2011)

18. Fahy, A.S., Clark, R.H., Glyde, E.F. & Smith, G.L. Vaccinia virus protein C16 acts intracellularly to modulate the host response and promote virulence. *J Gen Virol* 89(10), 2377–2387. https://doi.org/10.1099/vir.0.004895-0 (2008)

19. Benfield, C.T., Mansur, D.S., McCoy, L.E., Ferguson, B.J., Bahar, M.W., Oldring, A.P. et al. Mapping the ikappaB kinase beta (IKKbeta)-binding interface of the B14 protein, a vaccinia virus inhibitor of IKKbeta-mediated activation of nuclear factor kappaB. *Biol. Chem.* 286(23), 20727–2035 https://doi.org/10.1074/jbc.M111.231381 (2011)

20. Yuwen, H., Cox, J.H., Yewdell, J.W., Bennink, J.R. & Moss, B. Nuclear localization of a double-stranded RNA-binding protein encoded by the vaccinia virus E3L gene. *Virology* 195(2), 732–44 https://doi.org/10.1006/viro.1993.1424 (1993)

21. Senkevich, T.G., Koonin, E.V. & Moss, B. Vaccinia virus F16 protein, a predicted catalytically inactive member of the prokaryotic serine recombinase superfamily, is targeted to nucleoli. *Virology* 417(2), 334–342. https://doi.org/10.1016/j.virol.2011.06.017 (2011)

22. Ferguson, B.J., Benfield, C.T., Ren, H., Lee, V.H., Frazer, G.L., Smadova, P. et al. Vaccinia virus protein N2 is a nuclear IRF3 inhibitor that promotes virulence. *J Gen Virol* 94(9), 2070–2081 https://doi.org/10.1099/vir.0.054114-0 (2013)

23. Knipe, D. M. Nuclear sensing of viral DNA, epigenetic regulation of herpes simplex virus infection, and innate immunity. *Virology* 479–480, 153–159 https://doi.org/10.1016/j.virol.2015.02.009 (2015).

24. Tsai, K. & Cullen, B.R. Epigenetic and epitranscriptomic regulation of viral replication [published online ahead of print]. *Nat Rev Microbiol* https://doi.org/10.1038/s41579-020-0382-3 (2020)

25. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korfach, J. & Turner, S.W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7(6), 461–465. https://doi.org/10.1038/nmeth.1459 (2010)

26. Stoiber, M.H., Quick, J., Egan, R., Lee, J.E., Celniker, S.E., Neely, R., Loman, N., Pennacchio, L. & Brown, J.B. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv* 2017:094672. https://doi.org/10.1101/094672 (2017)

27. Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E. & Goull, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 21(1), 30 https://doi.org/10.1186/s13059-020-1935-5 (2020)

28. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. & Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci* 89(5), 1827–1831 https://doi.org/10.1073/pnas.89.5.1827 (1992)

29. Feederle, R. & Schepers, A. Antibodies specific for nucleic acid modifications. *RNA Biol* 14(9), 1089–1098. https://doi.org/10.18720/15476286.2017.1295905 (2017)

30. Müller, C.A., Boemo, M.A., Spingardi, P. et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods* 16, 429–436 https://doi.org/10.1038/s41592-019-0394-y (2019)
31. Nehme, Z., Pasquereau, S. & Herbein, G. Control of viral infections by epigenetic-targeted therapy. *Clin Epigenet* 11, 55 https://doi.org/10.1186/s13148-019-0654-9 (2019)

32. Kaaden, D. R., Walz, C.P. Czerny, U. Wernery, U. & Allen, W. R. Progress in the development of a camel pox vaccine. Proceeding of the 1st Int. Camel Conference, 47-49 (1992)

33. Saud, Z. & Butt, T.M. Another case of mistaken identity? Vaccinia virus in another live Camelpox vaccine. *Biologicals* 65, 39-41. https://doi.org/10.1016/j.biologicals.2020.04.002 (2020)

34. Maracci, M. Khalafla, A.I. Al Hammad, Z.M. Monaco, F. Cammà, C. Yusof, M.F. Al Yammahi, S.M. Mangone, I. Valleriani, F. Alhosani, M.A. Decaro, N. Lorusso, A. Almuhaier, S.S. Savini, G. Genome Sequencing of a Camelpox Vaccine Reveals Close Similarity to Modified Vaccinia virus Ankara (MVA). *Viruses* 12(8), E786. https://doi.org/10.3390/v12080786 (2020)

35. Lefkowitz, E.J., Upton, C., Changayil, S.S., Buck, C., Traktman, P. & Buller, R.M. Poxvirus Bioinformatics Resource Center: a comprehensive Poxviridae informational and analytical resource. *Nucleic Acids Res* 33, D311–316 https://doi.org/10.1093/nar/gki110 (2005)

36. Senkevich, T.G., Bruno, D., Martens, C., Porcella, S.F., Wolf, Y.I. & Moss, B. Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 112(35), 10908-10913. https://doi.org/10.1073/pnas.1514809112 (2015)

37. Esteban, M. & Holowczak, J.A. Replication of vaccinia DNA in mouse L cells. I. In vivo DNA synthesis. *Virology* 78(1), 57–75 https://doi.org/10.1002/6822(77)90078-2 (1977)

38. Pogo, B.G.T. & O’Shea, M.T. The mode of replication of vaccinia virus DNA. *Virology* 84(1), 1-8. https://doi.org/10.1016/0042-6822(78)90213-1 (1978)

39. Parkinson, J.E., Sanderson, C.M. & Smith, G.L. The vaccinia virus A38L gene product is a 33-kDa integral membrane glycoprotein. *Virology* 214(1), 177-188 https://doi.org/10.1006/viro.1995.9942 (1995)

40. Sanderson, C.M., Parkinson, J.E., Hollinshead, M. & Smith, G.L. Overexpression of the vaccinia virus A38L integral membrane protein promotes Ca2+ influx into infected cells. *J Virol* 70(2), 905-914 https://doi.org/10.1128/JVI.70.2.905-914.1996 (1996)

41. Alcamí, A., Symons, J.A. & Smith, G.L. The vaccinia virus soluble alpha/beta interferon (IFN) receptor binds to the cell surface and protects cells from the antiviral effects of IFN. *J Virol* 74(23), 11230-11239. https://doi.org/10.1128/jvi.74.23.11230-11239.2000 (2000)

42. Senkevich, T.G., Weisberg, A.S. & Moss, B. Vaccinia virus E10R protein is associated with the membranes of intracellular mature virions and has a role in morphogenesis. *Virology*, 278(1), 244-252 https://doi.org/10.1006/viro.2000.0656 (2000)

43. Price, N., Tscharke, D.C., Hollinshead, M. & Smith, G.L. Vaccinia virus gene B7R encodes an 18-kDa protein that is resident in the endoplasmic reticulum and affects virus virulence. *Virology* 267(1), 65-79 https://doi.org/10.1006/viro.1999.0116 (2000)

44. Meisinger-Henschel, C., Späh, M., Lukassen, S., et al. Introduction of the six major genomic deletions of modified vaccinia virus Ankara (MVA) into the parental vaccinia virus is not sufficient to reproduce an MVA-like phenotype in cell culture and in mice. *J Virol* 84(19) 9907-9919 https://doi.org/10.1128/JVI.00756-10 (2010)

45. Turner, P.C. & Moyer, R.W. The vaccinia virus fusion inhibitor proteins SPI-3 (K2) and HA (A56) expressed by infected cells reduce the entry of superinfecting virus. *Virology* 380(2), 226-233 https://doi.org/10.1016/j.viro.2008.07.020 (2008)

46. Roberts, K.L., Breiman, A., Carter, G.C., et al. Acidic residues in the membrane-proximal stalk region of vaccinia virus protein B5 are required for glycosaminoglycan-mediated disruption of the extracellular enveloped virus outer membrane. *J Gen Virol* 90(Pt 7), 1582-1591 https://doi.org/10.1099/vir.0.009092-0 (2009)

47. Sood, C.L. & Moss, B. Vaccinia virus A43R gene encodes an orthopoxvirus-specific late non-virion type-1 membrane protein that is dispensable for replication but enhances intradermal lesion formation. *Virology* 396(1), 160-168 https://doi.org/10.1016/j.viro.2009.10.025 (2010)

48. Owji, H., Nezafat, N., Negahdari-pour, M., Hajiebrahimi, A. & Ghasemi, Y. A comprehensive review of signal peptides: Structure, roles, and applications. *Eur J Cell Biol* 97(6), 422-441 https://doi.org/10.1016/j.ejcb.2018.06.003 (2018)

49. Duggan, A.T., Klunk, J., Porter, A.F. et al. The origins and genomic diversity of American Civil War Era smallpox vaccine strains. *Genome Biol* 21, 175 https://doi.org/10.1186/s13059-020-02079-z (2020)

50. Tombácz, D., Przasák, I., Szucs, A., Dénes, B., Snyder, M. & Boldogkoi, Z. Dynamic transcriptome profiling dataset of vaccinia virus obtained from long-read sequencing techniques. *Gigascience* 7(12), giy139 https://doi.org/10.1093/gigascience/giy139 (2018)

51. Tombácz, D., Przasák, I., Csabai, Z., et al. Long-read assays shed new light on the transcriptome complexity of a viral pathogen. *Sci Rep* 2020 10(1), 13822 https://doi.org/10.1038/s41598-020-70794-5 (2020)

52. De Coster, W., D’Hert, S., Schultz, D.T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15), 2666-2669. https://doi.org/10.1093/bioinformatics/bty149 (2018)

53. Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 37(5), 540-546. https://doi.org/10.1038/s41587-019-0072-8 (2019)
54. Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E. Bandage: interactive visualisation of de novo genome assemblies. *Bioinformatics* 31(20), 3350-3352. https://doi.org/10.1093/bioinformatics/btv383 (2015)

55. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18), 3094-3100 https://doi.org/10.1093/bioinformatics/btv383 (2018)

56. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(16), 2078-2079 https://doi.org/10.1093/bioinformatics/btp352 (2009)

57. Mikheenko, A., Bzikadze, A.V., Gurevich, A., Miga, K.H. & Pevzner, P.A. TandemTools: mapping long reads and assessing/improving assembly quality in extra-long tandem repeats. *Bioinformatics* 36(1), i75-i83 https://doi.org/10.1093/bioinformatics/btaa440 (2020)

58. Vaser, R., Sović, I., Nagarajan, N., Šikić, M.: Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5), 737-746 https://doi.org/10.1101/gr.214270.116 (2017)

59. Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villavecchia, J., Habermann, B., Akhtar, A. & Manke., T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications* 9(1), 189 https://doi.org/10.1038/s41467-017-02525-w (2018)

60. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. et al. Versatile and open software for comparing large genomes. *Genome Biology* 5(2), R12. https://doi.org/10.1186/gb-2004-5-2-r12 (2004)

61. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119 https://doi.org/10.1186/1471-2105-11-119 (2010)

62. Humann, J.L., Lee, T., Ficklin, S. & Main, D. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. *Methods Mol Biol* 1962, 29-51 https://doi.org/10.1007/978-1-4939-9173-0_3 (2019)

63. Jones, P., Binns, D., Chang, H.Y. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu03 (2014)

64. Camacho, C., Coulouris, G., Avagyan, V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421 https://doi.org/10.1186/1471-2105-10-421 (2009)

65. Almagro Armenos, J.J., Tsirigos, K.D., Sønderby, C.K. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019 37(4), 420-423 https://doi.org/10.1038/s41587-019-0036-z (2019)

66. Arias-Carrasco, R., Vásquez-Morán, Y., Nakaya, H.I. et al. StructRNAfinder: an automated pipeline and web server for RNA families prediction. *BMC Bioinformatics* 19, 55 https://doi.org/10.1186/s12859-018-2052-2 (2018)

67. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36 PMID: 7584402 (1994)

68. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol* 8(2), R24 https://doi.org/10.1186/gb-2007-8-2-r24 (2007)

### Tables

| Metric                  | Raw Reads | >3000 Viral DNA Read Set |
|-------------------------|-----------|-------------------------|
| Number of Reads         | 405,925   | 16,059                  |
| Cumulative Size         | 828,487,274 | 94,298,074            |
| Average Read Length (bp)| 2,041     | 5,872                   |
| N50 (bp)                | 6,507     | 6,174                   |

Table 1 Read metrics of sequencing before and after contamination removal.
| Protein | Amino Acid Sequence | SignalP v5.0 prediction | Uniprot prediction |
|--------|---------------------|------------------------|--------------------|
| A38L   | MSRVRISLYLYTLVIVTTKTIEYTACND | MSRVRISLYLYTLVIVTTKT | No signal peptide predicted |
|        | TIIIPCTIDNPTKRYIRKWKLNDHDLTYNKTTK | | |
|        | SKTTLILSWHTSARLHSLSLDSVSLIMEYDYNQLKDVINTDNDTTSFVKIYFH | | |
|        | KDILPGTYCTGDNIGKSTVQKVLQHTNW | | |
|        | FNDYQTMILMFIFGTIGTLFLFLEITYTSISVVFSTNLGLQVFGCIAEMELCAGFLYFPSMF | | |
|        | TLRHIIGLLMTLPSLIITKVFSFVLLCKS | | |
|        | SCAVHLIIYQLAGYIITVGLGLSLKECVDFGTLLSLGTIVMSEHFSSSLFLVCFPSTQQRDVY | | |
| C8L    | MSAIRFIACLYLISIFGNCEDPYQPFDDK | MSAIRFIACLYLISIFGNC | MSAIRFIACLYLISIFGNCHE |
|        | LNITLIDLITYELVLPYTVNTDTTSFVKIYF | | |
|        | KNFWITVMKWCPDFTSVYVTSIDHNLNIQFADTQEDKICTIDVKARC | | |
|        | KHLTKEVTQVQEAYRSLSDLSFCDSIDLEIDLENTSTNTTSTVLKSYELMLPKRACKSHN | | |
| HA     | MTRLPIILLLLLSLVYATFPQTSKIGGATTLSNRINNTNDYVVMASAYKEPSNILLAASKDVLÝFDNYTDKISKSYDPSYYDLVTTITIKSLTARD | MTRLPIILLLLLSLVYA | MTRLPIILLLLLSLVYA |
|        | AGTYVCAFMSTPTDNTKDCYEEIVSTELIVNTDESTIDIIISSTHSPETSSEPKPYIDNSNC | | |
|        | SSFVEIATPEPITNVEDHDGTVTDVSINTTVSASSGSTTDETPETDEHEDHTVTDVTYSYT | | |
|        | TVSTSSGIVTTSTTDDADLTDYNTNDTVPSTTVGGSSTTSINYKDFVIEIGITALIILSVAFIFCITYYINNLKRYERTKENTKV | | |
| B19R_0 | MKMTKMMVHIFVSSLLLLLFHSYADIENIEITEFFNKMRTDLPAKLDSWLNPCAMFGGTMNDIAALGEPSAKCPPIEDSLLSHRYKDYVVKWRLEKNRRQQVSNKRVKGDLWIANYTSKFKSNR | MKMTKMMVHIFVSSLLLLLFHSYA | MTKMVMHIFVSSLLLLLF |
|        | RYLCSTVTTKNGDCVQIVRSIKKPPSPICPKTEGTHDGYGILLCIGIGLYAKHYHNYITWYKDN | | |
|        | KEINIDDIKYSQTGGKLIHNPELEDGNYRNCYVHYDDVRKMM | | |
| E10R   | MNPKHWGAVWTTIIFIVLSQAGLGDIEACKRLYTVSTLPACRRHATIAEDNNVMSMDLNYYYFFIRLFFNLASDPKAYIDTVKVPNL | MNPKHWGAVWTTIIFIVLSQAGLGD | MNPKHWGR |
|        | WTDGGLNLLSKNYGIKTWQTMYTNVPEGTYDISAPKNDVFSFWVKFEOQGYKVEEYCTGPP | | |
|        | TVTLTEYDDHYPATRGGSKPIYKRGDMCDIY | | |
| B8R    | MRYIIILAVLFLINSIHA | MRYIIILAVLFLINSIHA | MRYIIILAVLFIN |
Table 2 Proteins containing predicted signal peptides. For each protein, the conventional signal peptide as stated by Uniprot is listed, as well as the signal peptide predicted by SignalP v5.0. A novel signal peptide was predicted by SignalP v5.0 for the protein A38L.

Figures

Bioinformatics pipeline used for the long-read only assembly of the Ducapox genome. Basecalling of reads was performed using Guppy v4.0.11. Adapter sequences were removed from the long-read sequences using Porechop v.0.2.4. Reads were subsequently filtered to a minimum length of 3000 bases using Nanofilt v2.6.0. An initial assembly was performed using Flye v.2.8, after which a BLAST search for each contig generated was performed against the NCBI nucleotide database. A file containing all non-viral reads was used to generate and exclusive viral reads set by mapping reads to the contamination contigs using Minimap2 v 2.17-r941, followed by extraction of the unmapped reads using Samtools v1.7. A Flye assembly was performed on the exclusive viral reads set, which was subsequently polished with TandeTools, followed by 3 rounds of Racon v.1.4.13 polishing, and a final polishing round using Medaka v0.11.5 to generate the final genome sequence (159,696 bp).
Figure 2

Read mapping coverage of genome assemblies for a. the initial Flye assembly and b. the final polished assembly. Read coverage was found to be more uniformly distributed in the final assembly in comparison to the initial assembly, which was found to have uneven read coverage distributions at the contig ends. This is indicative of the final polished assembled containing terminal repeat sequence lengths that more closely match that of the ground truth.

Figure 3

Dotplot comparison of the Ducapox long read assembly vs the closest matching viral genome, that of VACV Acambis 3000 MVA. Genomic deletions of 5449 bp and 916 bp in size are illustrated. The VACV Acambis 3000 MVA was also found to be 227 bp and 435 bp longer at its ends, with respect to the Ducapox genome.
Figure 4

Annotated Ducapox gene map. The genome contained a total of 191 genes.
AGAAGRC motif (n=17)

Figure 5
Statistical plot and sequence logo of the AGAAGRC motif. The statistical plot is based on 17 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model can be seen around the central AAG nucleotides. A Tomtom search of the motif detected no similar known motifs.

AARRRGATKH motif (n=42)

Figure 6
Statistical plot and sequence logo of the AARRRGATKH motif. The statistical plot is based on 42 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model can be seen around the central GA nucleotides. A Tomtom search of the motif showed the reverse-complement to most closely match MA0467.1 (Crx; Mus musculus) in the JASPAR database.
Figure 7
Statistical plot and sequence logo of the SANGGTM motif. The statistical plot is based on 40 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model can be seen around the central GT nucleotides. A Tomtom search of the motif showed the reverse-complement to most closely match MA0039.3 (KLF4; Homo sapiens) in the JASPAR database.

Figure 8
Statistical plot and sequence logo of the WWAATGWC motif. The statistical plot is based on 77 regions within the genome that contain the motif sequence. Signal fluctuation away from the canonical model can be seen around the central TGT nucleotides. A Tomtom search of the motif showed the reverse-complement to most closely match MA1112.1 (NR4A1; Homo sapiens) in the JASPAR database.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation1.docx
- SupplementaryInformation2SignalPSummary.docx
- SupplementaryInformation3DucapoxMatureProteins.fasta.docx