# The Pan-SL-CoV/GD sequences may be from contamination.

#### Daoyu Zhang.

# ABSTRACT

Recently, There were much hype about an alleged SARS-like coronavirus being found in samples of Malayan pangolins (Manis Javanica) possessing nearly identical RBD to the SARS-CoV-2 coronavirus. Prominent journals cite the alleged discovery to claim that pangolins may be one of a possible intermediate host for the zoonotic transmission of SARS-CoV-2 to humans.

Here, we report that all databases used to support such a claim, upon which metagenomic analysis was possible, contained unexpected reads and was in serious risk of contamination. Here we also report that the presence of unexpected reads are directly related to the presence of coronavirus reads. Finally, we deduced the actual causative agent of the death of the pangolins sampled in GuangDong 2019 where the claim of coronavirus detections was made.

# **METHODS**

## The NCBI Trace tool

The NCBI SRA archive come with it's own tool called Trace, which identifies the origin or reads within the SRA dataset through the recognition of unique K-mers within the nucleotide sequence. Multiple reads of 32 nucleotides is taken from each read to identify the reads toward an origin by comparison with a large database of reference sequences, which produces a classification signal. Then read of 64 nucleotides are taken from each of the read for definitive mapping toward species in the reference database. If any one of the 32nt or 64nt K-mers are found in more than one reference sequence, the reads are instead classified at the lowest phylogenetic classification node where reference sequences containing such a K-mer is found.

The 32nt TRACE generate a "strong signal" classification of sequence origin useful for the deduction of the content of the sample by organism of origin, accessed via the NCBI Krona charting tool,

While the 64nt TRACE generate a definitive classification signal used for the exact tracing of reads to the origin from a specific Species/Taxon, used for the exact classification of reads.

Both the 32nt and 64nt TRACE analysis classify their reads according to the lowest common taxonomical node where K-mers from said read are present in the reference sequence database, a strategy known as "lowest non-ambiguous mapping". Such a strategy avoids the problem with RNA degradation or sequencing errors by excluding potential errors in reads, without introducing potential ambiguous classification by clustering ambiguous reads under the lowest common classification node such ambiguity is found.

Therefore, if TRACE gives an identification to a specific taxonomical node for a sequence read, it could be from any of the taxonomical nodes and species classified under the node, but it could not be from a taxonomical node or species that is not under said node. E.g. if TRACE says hominoidea which was classified under Catarrhini; Simiiformes; Haplorrhini; Primates; Euarchontoglires, Then it can't be from a pangolin since pangolins (Manis Spp.) are classified under Pholidota; Laurasiatheria. The lowest common classification node between Primates and Pangolins is Boreoeutheria—reads from parts of the genomes shared between Primates and Pangolins will only be classified to Boreoeutheria, but not further classified down toward either Laurasiatheria or Euarchontoglires. And definitely will not be classified individually toward Pholidota or Primates, or any child nodes or phylogenetic nodes under them.

# **Specific BLAST analysis**

Whenever a genus or species is provided by analysis, a specific BLAST analysis is performed to confirm the presence of reads toward the exact species by a search of the database in question with representative reference sequences of the specific species in question in look for matches that is either: 100% match, or: contained no 100% matches on BLAST when queried against the Pangolin reference sequences available on GanBank.

# RESULTS

The Accession numbers and contents of all Pan-SL-CoV/GD related sequencing experiments are listed under the following table.

Table 1: List of available GD Pangolin sample datasets as provided under NCBI GenBank. By Accession number, size and citation by thesis (if claimed to have SARS-CoV-2 related reads by paper).

Accession number	Size	SARS-CoV-2-like Coronavirus
		Identified and Cited?
<u>SRX6893158</u>	16,491,648	
<u>SRX6893157</u>	9,275,501	Lung12 [3] SRR10168374
<u>SRX6893156</u>	22,220,187	Lung11 [1]
<u>SRX6893155</u>	18,067,615	Lung09 [1] [3] SRR10168376
<u>SRX6893154</u>	16,414,925	Lung08 [1] [3] [4]
		SRR10168377
<u>SRX6893153</u>	19,045,923	Lung07 [1] [3] [4]
		SRR10168378
<u>SRX6893152</u>	13,527,964	
<u>SRX6893151</u>	16,068,654	
<u>SRX6893150</u>	12,967,281	
<u>SRX6893149</u>	12,590,769	
<u>SRX6893148</u>	15,273,939	

<u>SRX6893147</u>	15,975,904	
<u>SRX6893146</u>	19,038,817	
<u>SRX6893145</u>	19,055,973	
<u>SRX6893144</u>	15,350,468	
<u>SRX6893143</u>	11,527,782	
<u>SRX6893142</u>	20,045,443	
<u>SRX6893141</u>	18,903,834	
<u>SRX6893140</u>	19,986,780	
<u>SRX6893139</u>	39,738,679	Lung02 [3] SRR10168392
<u>SRX6893138</u>	22,900,426	
<u>SRX7756769</u>	107,267,359 PRJNA607174**	M1[2]***
<u>SRX7756766</u>	273,651,431 PRJNA607174**	
<u>SRX7756765</u>	196,761,202 PRJNA607174**	
<u>SRX7756764</u>	222,286,763 PRJNA607174**	
<u>SRX7756763</u>	212,161,250 PRJNA607174**	
<u>SRX7756762</u>	232,433,120 PRJNA607174**	M6[2]***
<u>SRX7756761</u>	113,900,941 PRJNA607174**	
SRX7732094	2,633*	"P2S"[3]

\*: "Design: This dataset contains coronavirus-like sequence reads, based on BLAST search."

\*\*: All available SRA datasets from PRJNA607174

\*\*\*:Actual SRA datasets identified from the "Extended Data Table 3" of [2]

# Article

Extended Data Table 3 | Identification of SARSr-CoV sequence reads in metagenomes from the lung of pangolins using the SARS-CoV-2 sequence (GenBank accession No. MN908947) as the reference

-				
_	No. mapped	Total reads*	Animal species	Sample ID
RX7756769 "pangolin 9"	496 ←9	107,267,359	Malayan pangolin	M1
	302	38,091,846	Malayan pangolin	M2
	14	79,477,358	Malayan pangolin	МЗ
lot available	1,100	32,829,850	Malayan pangolin	M4
	56	547,302,862	Malayan pangolin	M5
RX7756762 "pangolin 2"	10 ←	232,433,120	Malayan pangolin	M6
	12	44,440,374	Malayan pangolin	M8
ot available	0	227,801,882	Malayan pangolin	M10
	o	444,573,526	Chinese pangolin	Z1

Fig.1 the "Extended Data Table 3" of [2]. SRA datasets identified in the available database is pointed out by an arrow, while SRA "runs" that failed to be identified in known datasets are outlined in a red square.

# Analysis of reads from The Available datasets using NCBI Trace.

		0	
Accession number and	Primary Mammalian	Primate-related results	Identification of
registration date	Trace results and	in Krona and read size	"Coronaviridae"
	percentage	by Кbp	as by Trace and
			total read size
SRX6893158	Manis javanica: 14.66%	N/D	N/D
20-Sep-2019			
SRX6893157	Boreoeutheria: 1.24%	Catarrhini 644546	N/D***
20-Sep-2019			
<u>SRX6893156</u>	Manis javanica: 7.51%	Homo sapiens 81948	Pangolin
20-Sep-2019	Homo sapiens: <b>0.03%</b>		coronavirus 2Kbp
<u>SRX6893155</u>	Homo sapiens: <b>0.37%</b>	Homininae 3534150	Pangolin
20-Sep-2019			coronavirus 5Kbp
<u>SRX6893154</u>	Homo sapiens: <b>0.02%</b>	Hominoidea 356003	Pangolin
20-Sep-2019			coronavirus
			154Kbp
SRX6893153	Homo sapiens: <b>0.01%</b>	Homo sapiens 162180	Pangolin
20-Sep-2019			coronavirus
			41Kbp
<u>SRX6893152</u>	Manis javanica: <b>2.87%</b>	N/D	N/D
20-Sep-2019	Euarchontoglires: 1.37%		
<u>SRX6893151</u>	Manis javanica: 7.47%	N/D	N/D
20-Sep-2019			
<u>SRX6893150</u>	Boreoeutheria: 1.91%	N/D	N/D
20-Sep-2019			
SRX6893149	Manis javanica: <b>1%</b>	Simiiformes 313069	N/D
20-Sep-2019			
<u>SRX6893148</u>	Manis javanica: <b>0.4%</b>	Catarrhini 194320	N/D
20-Sep-2019			
SRX6893147	Manis javanica: 2.71%	Catarrhini 69937	N/D
20-Sep-2019			
<u>SRX6893146</u>	Boreoeutheria: 1.72%	Hominoidea 231755	N/D
20-Sep-2019			
<u>SRX6893145</u>	Homininae: 0.27%	Homininae 2536765	N/D
20-Sep-2019	Manis javanica: <b>1.01%</b>		
SRX6893144	Manis javanica: 0.62%	Hominoidea 166628	N/D
20-Sep-2019			
SRX6893143	Manis javanica: 1.63%	N/D	N/D
20-Sep-2019			
SRX6893142	Manis javanica: 1.28%	Simiiformes 57084	N/D

Table 2. The Trace result of Known GD Pangolin datasets when examined using NCBI Trace SRA.

20-Sep-2019			
SRX6893141	Boreoeutheria: 1.41%	N/D	N/D
20-Sep-2019			
SRX6893140	Boreoeutheria: 1.56%	N/D	N/D
20-Sep-2019			
<u>SRX6893139</u>	Homo sapiens: <b>0.01%</b>	Homo sapiens 491120	Pangolin
20-Sep-2019			coronavirus 2Kbp
<u>SRX6893138</u>	Boreoeutheria: 1.67%	Homininae 2761176	N/D
20-Sep-2019			
<u>SRX7756769</u>	Homo sapiens: 0.03%	Homo sapiens 5457929	Bat SARS-like
18-Feb-2020			coronavirus 2Kbp
			Wuhan seafood
			market
			pneumonia virus
			2Kbp
<u>SRX7756766</u>	Manis javanica: 78.6%	Cercopithecidae 3116	Betacoronavirus
18-Feb-2020			2Kbp**
<u>SRX7756765</u>	Manis javanica: 87.17%	Cercopithecinae 11339	N/D****
18-Feb-2020			
<u>SRX7756764</u>	Manis javanica: 48.39%	Cercopithecidae 22600	N/D
18-Feb-2020			
<u>SRX7756763</u>	Manis javanica: <b>94.95%</b>	Cercopithecidae 5076	N/D
18-Feb-2020			
<u>SRX7756762</u>	Manis javanica: 95.37%	Catarrhini* 2831	Nidovirales OKbp
18-Feb-2020			
SRX7756761	Manis javanica: 13.63%	Chlorocebus sabaeus	N/D
18-Feb-2020		498506	
SRX7732094	N/A***	N/A	Pangolin
15-Feb-2020			coronavirus***

\*: Chlorocebus Sabaeus

\*\*:Not claimed as being SARS-CoV-2 related in the original publication. Likely unrelated.

\*\*\*Not analyzable. All Non-Coronavirus data filtered out. Leaving only 2,633 reads, all of which can be mapped to the SARS-CoV-2 reference genome.

\*\*\*\*8 reads as claimed by [10]

#### **Specific BLAST analysis**

In order to determine the authenticity of the Primate-related reads in the datasets, Specific BLAST analysis is carried out for all datasets that possessed claimed or analyzed reads of coronaviridae-related viruses. An 100% full-length match that does not map to non-primates confirms Authenticity of read.

	select all 100 sequences selected					Grap	hics Distance tree of results
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	<u>SRX7756762</u>	279	1047	0%	4e-68	100.00%	SRA:SRR11119766.160125840.2
	<u>SRX7756762</u>	279	1366	0%	4e-68	100.00%	SRA:SRR11119766.138036805.1
	<u>SRX7756762</u>	279	967	0%	4e-68	100.00%	SRA:SRR11119766.101239747.1
	<u>SRX7756762</u>	279	1624	0%	4e-68	100.00%	SRA:SRR11119766.46413326.2
Chlo	rocebus sabaeus isolate 1994-021 unplaced genomic sca						
dna							
1339	488						
<u>Dista</u>	ance tree of results MSA viewer 🔞						

Fig.2a Specific BLAST analysis on the PRJNA607174 dataset, <u>SRX7756762</u>, that contained claimed SARS-CoV-2 related coronavirus reads. The 100% full-length matches clearly indicate presence of Primate-derived material.

	select all 100 sequences selected		Gen	<u>Bank</u>	<u>Grap</u>	hics	Distance	ree of results
	Description		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Macaca mulatta isolate Rh22777_5890-1b major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332833.1
	Macaca mulatta isolate Rh22335_5775-3 major histocompatibility_complex_genomic_sequence		279	279	100%	2e-71	100.00%	KT332608.1
	Macaca mulatta isolate Rh22335_5725-2 major histocompatibility_complex_genomic_sequence		279	279	100%	2e-71	100.00%	KT332521.1
	Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332463.1
>gi TAA AGA CAA	IIIJSKAJSKILLISJOG.JOUZSAUZ JOUZSAUZ (DUUZSAU ITCCTTGGGAATCGCCATATGGGATGGCTGGGTCGACATAGGTACATCTAGTTCT ITCCTTGAGGAATCGCCATACTGTTTTCCATAATGGTTGAACTAGTTTACAATCCCAC ACAGTGTAAAAGTGTTCCTATTTCTCCAC							
~	Uncultured organism clone VC1C968TR genomic sequence	230	230	1	00% 2	e-56	94.04%	<u>GQ879596.1</u>
<b>_</b>	Synthetic construct Pan troglodytes LINE-1L 1Pt retrotransposon tagged with EGFP sequence	219	219	1	00% 4	le-53	92.72%	KF661301.1
~	Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YAC/BAC#26-2	219	219	1	00% 4	le-53	92.72%	AB553834.1
~	Mus musculus NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone;F630221F08 productum	204	204	8	6% 1	e-48	94.66%	AK171052.1
~	Mus musculus bone marrow macrophage cDNA, RIKEN full-length enriched library, clone:G530008A19 product.hypothetical pi	204	204	8	6% 1	e-48	94.66%	AK149653.1
~	Ralstonia solanacearum genome assembly 9 genomes, chromosome : V	202	202	1	00% 4	le-48	90.73%	LN899823.1
~	Canis lupus familiaris breed Labrador retriever chromosome 06a	154	3044	4 9	8% 1	e-33	86.43%	CP050586.1
~	Canis lupus familiaris breed Labrador retriever chromosome 04a	154	4569	9 1	00% 1	e-33	85.23%	CP050572.1
~	Canis lupus familiaris breed Labrador retriever chromosome 06b	154	3042	2 9	8% 1	e-33	86.43%	CP050622.1

Fig.2b BLAST result on the returned sequence revealed it as a Primate-derived MHC complex gene that is not found in non-primates, confirming Primate origin.

<b>~</b>	select all 100	sequences selected					Grap	hics Distance tree of results
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.269072261.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255768440.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255768440.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255318754.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.254520929.1
	SRX7756766		279	6344	0%	5e-67	100.00%	SRA:SRR11119762.251645135.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.234036838.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.211208832.2
	SRX7756766		279	9108	0%	5e-67	100.00%	SRA:SRR11119762.199583624.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.198110623.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.196936636.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.196936636.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.133631622.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.108819247.2
De	scription	Macaca mulatta isolate AG07107 chromosome 3 genomic sca $\dots$						
Мо	lecule type	dna						
Qu	ery Length	17855752						
Ot	her reports	Distance tree of results MSA viewer						

Fig.3a Specific BLAST analysis of <u>SRX7756766</u> revealed large amount of 100% full-length matches with Macaca Mulatta.

	Macaca mulatta Major Histocompatibility Complex BAC MMU370002, complete sequence	279	2	279	100%	2e-71	100.00%	AC148706.1
	Macaca mulatta Major Histocompatibility Complex BAC MMU122H23, complete sequence	279	5	521	100%	2e-71	100.00%	AC148677.1
~	Macaca mulatta Major Histocompatibility Complex BAC MMU065H09, complete sequence	279	2	279	100%	2e-71	100.00%	AC148671.1
	Macaca mulatta Major Histocompatibility Complex BAC MMU038L02, complete sequence	279	2	279	100%	2e-71	100.00%	AC148668.1
~	Papio anubis clone rp41-22m16, complete sequence	279	5	54	100%	2e-71	100.00%	AC113268.8
	Papio anubis clone rp41-280n2, complete sequence	279	4	165	100%	2e-71	100.00%	AC091778.13
~	Papio anubis clone rp41-5m22, complete sequence	279	2	279	100%	2e-71	100.00%	AC136143.4
	Papio anubis clone rp41-192i11, complete sequence	279	5	526	100%	2e-71	100.00%	AC091671.28
~	Macaca mulatta chromosome 9 CH250-18D2, complete sequence	279	5	548	100%	2e-71	100.00%	CT573219.3
	Rhesus Macaque CHR4 BAC CH250-23P16 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Librar	279	2	279	100%	2e-71	100.00%	AC169807.2
~	Rhesus Macaque CHR4 BAC CH250-476F18 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libra	279	2	279	100%	2e-71	100.00%	AC171646.5
	Chlorocebus aethiops BAC clone CH252-163P9 from chromosome 5, complete sequence	278	5	56	99%	7e-71	100.00%	AC239684.4
	MACACA MULATTA BAC clone CH250-192J17 from chromosome unknown, complete sequence	278	5	52	100%	7e-71	100.00%	AC215693.3
	Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence	274	2	274	100%	9e-70	99.34%	KT332463.1
	Macaca mulatta isolate Rh9_6570-3 major histocompatibility complex genomic sequence	274	2	274	100%	9e-70	99.34%	<u>KT331777.1</u>
~	Macaca mulatta isolate Rh9_6550-1b major histocompatibility complex genomic sequence	274	2	274	100%	9e-70	99.34%	KT331733.1
	Macaca mulatta isolate Rh9_6526-2 major histocompatibility complex genomic sequence	274	2	274	100%	9e-70	99.34%	KT331675.1
~	Macaca mulatta isolate Rh18665_5547-1b major histocompatibility complex genomic sequence	274	2	274	100%	9e-70	99.34%	KT329509.1
~	Macaca mulatta isolate Rh23717 clone 4777 major histocompatibility complex-B genomic sequence	274	2	274	100%	9e-70	99.34%	KJ913523.1
~	Macaca mulatta isolate Rh23108 clone 4769-2 major histocompatibility complex-B genomic sequence	274	2	274	100%	9e-70	99.34%	KJ913420.1
	Eukaryotic synthetic construct chromosome 18	:	202	4103	969	6 4e-4	18 91.789	6 <u>CP034496.1</u>
	Eukaryotic synthetic construct chromosome 19		196	2020	969	6 2e-4	6 92.099	6 <u>CP034522.1</u>
	Eukaryotic synthetic construct chromosome 19		196	2130	969	6 2e-4	16 92.099	6 <u>CP034497.1</u>
-	Eukaryotic synthetic construct chromosome 16		196	3654	969	6 2e-4	6 91.619	6 <u>CP034494.1</u>
<b>-</b>	Eukaryotic synthetic construct chromosome 15		196	4994	979	6 2e-4	6 92.099	6 <u>CP034493.1</u>
	Eukaryotic synthetic construct chromosome 14		196	773	969	6 2e-4	16 92.099	6 <u>CP034492.1</u>
<b>_</b>	Eukaryotic synthetic construct chromosome 13		191	7988	979	6 9e-4	15 90.91%	6 <u>CP034516.1</u>
	Eukaryotic synthetic construct chromosome Y		191	2781	969	6 9e-4	15 91.379	6 <u>CP034510.1</u>
	Eukaryotic synthetic construct chromosome 20		191	4517	969	6 9e-4	15 91.379	6 <u>CP034499.1</u>
-	Eukaryotic synthetic construct chromosome 13		191	7988	979	6 9e-4	15 90.91%	6 <u>CP034491.1</u>
-	Eukaryotic synthetic construct chromosome 21		185	1053	969	6 4e-4	13 90.65%	6 <u>CP034500.1</u>
-	Eukaryotic synthetic construct chromosome 17		185	2450	969	6 4e-4	13 90.65%	6 <u>CP034495.1</u>
-	Eukaryotic synthetic construct chromosome 22		183	933	969	6 2e-4	12 90.58%	6 <u>CP034501.1</u>

Fig.3b BLASTing such matches gives 1005 matches to only Primates, and with no matches outside of Primates. This indicate that <u>SRX7756766</u> also contained significant amount of material derived from primates.

Results for	2:lcl Query_13045 gnl SRA SRR11119762.182596220.2 182596220 (Biologi ♥	016011311 0119 10p 20 Will 0	ррси					
Program	BLASTN 🕄 Citation 🗸	Type common name, bin	omial,	taxid or (	group n	ame		
Database	nt See details 🗸	+ Add organism						
Query ID	lcl/Query_13045	Percent Identity	E value	,		01	ierv Cov	erage
Description	gnl SRA SRR11119762.182596220.2 182596220 (Biological)	to		to		٦È		-o
Aolecule type	dna						`	
Query Length	151						Filter	Reset
Other reports	Distance tree of results							
Descriptions	Graphic Summary Alignments Taxonomy							
Sequences p	producing significant alignments	Download 🗡	Ма	inage Co	lumns	~ s	how 1	000 🗸 📀
select all	0 sequences selected				nk <u>Gr</u>			
	Description		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	D: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA		279	279	100%	2e-71	100.00%	XM_015150059.2
	D: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA		279	279	100%	2e-71	100.00%	XM_015150058.2
PREDICTEI	<u>D: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA</u>		279	279	100%	2e-71	100.00%	XM_015150057.2
Papio anubi	is anubis NIPSNAP1 protein (NIPSNAP1) gene, partial cds; and merlin (NF2) and S-100	//CaBP-type calcium binding domain	279	5181	100%	2e-71	100.00%	AH012454.2
PREDICTEI	<u>D: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA</u>		279	279	100%	2e-71	100.00%	XM_015457315.1
	<u>D: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA</u>		279	279	100%	2e-71	100.00%	XM_005567555.2
	<u>D: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA</u>		279	279	100%	2e-71	100.00%	XM_005567554.2
Chlorocebu	s aethiops BAC clone CH252-146I12 from chromosome 6, complete sequence		279	3318	100%	2e-71	100.00%	AC241602.2
Chlorocebu	s aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence		279	681	100%	2e-71	100.00%	AC239463.3
Macaca mu	latta BAC CH250-74N24 (Children's Hospital Oakland Research Institute Rhesus maca	que Adult Male BAC Library) complet	279	6774	100%	2e-71	100.00%	AC204493.6
Chlorocebu	s aethiops BAC clone CH252-257N12 from chromosome 6, complete sequence		274	6168	100%	9e-70	99.34%	AC241575.3
Chlorocebu	s aethiops BAC clone CH252-518J9 from chromosome 6, complete sequence		274	4718	100%	9e-70	99.34%	AC241458.2
Chlorocebu	s aethiops BAC clone CH252-124F12 from chromosome 6, complete sequence		274	2252	100%	9e-70	99.34%	AC241509.3
Chlorocebu	s aethiops BAC clone CH252-175F15 from chromosome 6, complete sequence		274	4253	100%	9e-70	99.34%	AC241845.3
Chlorocebu	s aethiops BAC clone CH252-57O21 from chromosome 6, complete sequence		274	2432	100%	9e-70	99.34%	AC241469.3

Fig.3C Presence of Primate-derived mRNA reads in <u>SRX7756766</u> confirms the Primate origin of these reads.

✓	select all 100 sequences selected					Grap	hics Distance tree of results
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.99831231.2
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.99831231.1
	SRX7756769	278	4814	1%	9e-69	100.00%	SRA:SRR11119759.88019245.2
	SRX7756769	278	5178	2%	9e-69	100.00%	SRA:SRR11119759.82130976.2
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.70689253.2
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.70689253.1
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.57405658.2
	SRX7756769	278	278	0%	9e-69	100.00%	SRA:SRR11119759.57405658.1
AC	073210.8						
н	omo sapiens BAC clone RP11-460N20 from 7, complete seq						
nı	cleic acid						
20	3396						

Fig.4a Similarly, <u>SRX7756769</u> contained large amount of reads that are 100% full-length matches to Human genomic DNA.

select all o	sequences selected			nk <u>G</u>			
	Description M	lax	Total	Query	E	Per.	Accession
Homo sanier	is chromosome 22 clone ABC11 000047178300 E22 complete sequence 2	78	456	100%	6e-71	100.00%	AC279316.1
Homo sapier	is actin related protein 2 pseudogene (LOC284441) on chromosome 19 2	78	278	100%	6e-71	100.00%	NG 022927.2
Homo sapier	Is TBC1 domain containing kinase (TBCK), RefSeqGene on chromosome 4 2	78	2140	100%	6e-71	100.00%	NG 034057.3
Homo sapier	is chromosome 15 clone VMRC59-280106, complete seguence 2	78	2291	100%	6e-71	100.00%	AC279072.1
Homo sapier	is chromosome 2 clone VMRC59-389K09, complete sequence 2	78	3905	100%	6e-71	100.00%	AC279037.1
Homo sapier	is chromosome 15 clone VMRC59-359A02, complete sequence 2	78	3589	100%	6e-71	100.00%	AC278991.1
Homo sapier	is chromosome 16 clone VMRC59-453B14, complete sequence 2	78	2239	100%	6e-71	100.00%	AC278975.1
Description	gnl SRA SRR11119759.88019245.2 88019245 (Biological)						
Molecule type	dna						
Ouery Length	150						
Other reports	Distance tree of results MSA viewer 😧						
	·						
Query ID	Icl Query_61480 Percent Identity E	value	9		(	Query Co	verage
Description	gnl SRA SRR11119759.70689253.170689253 (Biological)		to				to
Molecule type	dna						
Query Length	150					Filter	Reset
Other reports	Distance tree of results 😧						
Descriptions	Graphic Summary Alignments Taxonomy						
Sequences	producing significant alignments Download $$	м	lanage	Colum	ns ~	Show	1000 🗸
Sequences	producing significant alignments Download ~ 41 sequences selected	м	lanage <u>Gen</u> B	<b>Colum</b>	i <b>ns</b> ~ <u>Graphic</u>	Show Show	1000 🗸
Sequences	producing significant alignments Download ~ 41 sequences selected	Max	Ianage GenB Total	Colum ank Query	ins ~ <u>Graphic</u> E	Show Show Show Show Show Show Show Show Show Show Show Show Show Show	1000 V
Sequences	producing significant alignments Download ~ 41 sequences selected Description	Max Score	GenB Total Score	Colum ank Query Cover	Graphic E value	Show S	1000 V
Sequences Select all PREDIC	producing significant alignments Download ~ 41 sequences selected Description TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA plane endogenous retrovirus group 3 applied 1. envelope (ERV3-1) trapectitivation 1. mPNA	Max Score 278	GenB Total Score 278	Colum ank Query Cover 100%	Graphic Graphic E value 7e-71 7e-71	Show Show	1000 ✔ ( nce tree of result Accession XM_034964253 NM_001007253
Sequences Select all PREDIC Homo si PREDIC	producing significant alignments Download  41 sequences selected  TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA plens endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 1. mRNA  TED: Pan boloddes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 2. mRNA	Max Score 278 278 278	GenE GenE Total Score 278 278 278	Colum ank Query Cover 100% 100%	Graphic Graphic E Value 7e-71 7e-71 7e-71	Show <u>Dista</u> Per. Ident 100.00% 100.00%	1000 ✓ nce tree of resu Accession XM_034964253 NM_001007253 XM_024357564
Sequences Select all PREDIC Homo sa PREDIC PREDIC	producing significant alignments Download  41 sequences selected  TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA pplens endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 1. mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA	Max Score 278 278 278 278	GenE Total Score 278 278 278 278 278	Colum ank Query Cover 100% 100% 100%	Graphic Graphic Value 7e-71 7e-71 7e-71 7e-71	Show [ <u>s Dista</u> Per. Ident 100.00% 100.00% 100.00%	1000 ♥ nce tree of resu Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774
Sequences Select all PREDIC Homo si PREDIC PREDIC Synthetic	producing significant alignments Download  41 sequences selected  TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA polens endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 1. mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12. mRNA	Max Score 278 278 278 278 278 278 278	GenE Total Score 278 278 278 278 278 278 278	Colum ank Query Cover 100% 100% 100% 100%	Graphic Graphic Value 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71	Show S Dista Per. Ident 100.00% 100.00% 100.00% 100.00%	1000 ♥ nce tree of resu Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774 AB590999.1
Sequences Select all PREDIC Homo si PREDIC Synthetic Homo si	producing significant alignments Download  Usequences selected  Description  TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1), mRNA pielena endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1), transcript variant 12, mRNA TED: Pan troglodytes endogenous retrovirus group 3 member 1. en	Max Score 278 278 278 278 278 278 278 278 278	GenE GenE Total Score 278 278 278 278 278 278 278 278 278 278	Colum ank Query Cover 100% 100% 100% 100%	Graphic           Graphic           Value           7e-71	Show Show Per. Ident 100.00%	1000 ✔ nce tree of resu Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774 AB590999.1 AK295189.1
Sequences  Sequences  Select all  PREDIC  PREDIC  Synthetic  Pan tog  Pan tog	producing significant alignments Download  Usescription  TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA polens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 1, mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 1, mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 1, mRNA TED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 1, mRNA construct DNA. cline: pEN21AE1261. Homo sapiens ERV3 gene for endogenous retrovirus accestral Env polyprotein precursor odytes BAC done CH251-623C19 from chromosome 7, complete sequence	Max Score 278 278 278 278 278 278 278 278 278 278	GenE GenE Total Score 278 278 278 278 278 278 278 278 278 278	Colum ank Query Cover 100% 100% 100% 100% 100%	Graphic Graphic Value 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71	Show Show Per. Ident 100.00% 100.00% 100.00% 100.00% 100.00% 100.00%	1000 ✔ nce tree of resu Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774 AB590999.1 AK295189.1 AC284799.2
Sequences  Select all  PREDIC  PREDIC  Synthetic  Pan tog  Pan tog  Pan tog	producing significant alignments Download  Usescription  TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA polens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA construct DNA, cline, pEN21AE1261, Homo sapiens ERV3 gene for endogenous retroviral sequence 3, without stop cod pielens cDNA FLI60255 complete ds. highly similar to HERV-R, 7q212, provirus ancestral Env polyprotein precursor odvikes BAC done CH251-565C10 from chromosome 7, complete sequence	Max Score 278 278 278 278 278 278 278 278 278 278	Gene Total Score 278 278 278 278 278 278 278 278 278 278	Colum ank Query Cover 100% 100% 100% 100% 100% 100%	Graphic Graphic Value 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71 7e-71	Show         Dista           Per.         Ident           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%	1000 ✔ nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 014956774 AB5909991 AC2951891 AC2951891 AC2947992 AC1483133
Sequences Sequences Select all PREDIC Homo si PREDIC Synthetic Homo si Pan tog Pan tog Pan tog Pan tog	producing significant alignments Download  Usescription  TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA polena endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA construct DNA. cline: pEN21AE1261. Homo sapiens ERV3 gene for endogenous retrovirus accestral Env polyprotein precursor dotkes BAC done CH251-565C10 from chromosome 7, complete sequence addres endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA	Max Score 278 278 278 278 278 278 278 278 278 278	GenB Total Score 278 278 278 278 278 278 278 278 278 278	Colum ank Query Cover 100% 100% 100% 100% 100% 100%	Graphic           E           value           7e-71	Show         Distance           Per.         Ident           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%	1000 ✓ nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AB5909991 AC1847992 AC1847992 AC148313.3 NM 001135586
Sequences Select all PREDIC PREDIC PREDIC Synthetic Homo si Pan tog Pan tog Pan tog Homo si	producing significant alignments Download  Usescription  TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA polens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodries endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA construct DNA cline; pEN21AE1261. Homo sapiens ERV3 gene for endogenous retrovirus accestral Env polyprotein precursor odvikes BAC done; CH251-523C19 from chromosome 7, complete sequence advikes BAC done; CH251-565C10 from chromosome 7, complete sequence advikes BAC done; CH251-565C10 from chromosome 7, complete sequence advikes BAC done; RP11-460N20 from 7, complete sequence	Max Score 278 278 278 278 278 278 278 278 278 278	GenB Constant	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100%	Graphic           Graphic           Value           7e-71	Show         Dista           Per.         Ident           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%	1000 ✔ nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AB590999.1 AC184799.2 AC148313.3 NM 001135588 AC073210.8
Sequences Select all PREDIC PREDIC PREDIC Synthetic Homo si Pan tog Pan tog Pan tog Homo si Homo si Homo si	producing significant alignments Download  Usescription  TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA poiens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA TED: Pan trodiodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 11, mRNA construct DNA, cline; pEN21AE 1261, Homo sapiens ERV3 gene for endogenous retrovirus accestral sequence 3, without stop cod pielens cDNA FLI60255 complete ds. highly similar to HERV-R, 7q212, provirus ancestral Env polyprotein precursor odvies BAC done CH251-65C10 from chromosome 7, complete sequence advies endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA pielens BAC done RP11-460N20 from 7, complete sequence pielens cDNA FLI23884 fis, clone LNG13819, highly similar to Human endogenous retrovirus ERV3, pol-env-31,TR region	Max Score 278 278 278 278 278 278 278 278 278 278	Landard Control Contro	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100%	Graphic           E         Value           7e-71         7e-71	Show         Dista           Per.         Ident           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%	1000 ✔ nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AB590999.1 AC184799.2 AC184799.2 AC148313.3 NM 001135588 AC073210.8 AK074464.1
Sequences Select all PREDIC PREDIC PREDIC Synthetic Homo si Pantog Pantog Pantog Homo si Homo si Homo si Homo si Homo si	producing significant alignments         Download           41 sequences selected         Description           TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         selens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 1, mRNA           TED: Pan baniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         TED: Pan troglodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           TED: Pan troglodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         Construct DNA, client, pRN21AE 1261, Homo sapiens ERV3 gene for endogenous retrovirus accestral sequence 3, without stop cod gelens cDNA FLI69255 complete cds, highly similar to HERVR, 7a212 provirus ancestral Env polyprotein precursor odytes BAC done CH251-565C10 from chromosome 7, complete sequence           codytes BAC done CH251-565C10 from chromosome 7, complete sequence         addres endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           plens BAC done RP11-450N20 from 7, complete sequence         addres endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           plens cDNA FLI23884 fis, clone LING13819, highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region         mdogenous retrovirus ERV3, pol-env-31TR region	Max Score 278 278 278 278 278 278 278 278 278 278	Gene           Total           Score           278	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	Graphic           E           value           7e-71	Show [ Dista Per. Ident 100.00% 100.0	1000 ✔ 0 nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AE590999.1 AC184799.2 AC184799.2 AC148313.3 NM 001135588 AC073210.8 AK074464.1 M12140.1
Sequences Select all PREDIC Homo si PREDIC PREDIC Synthetic Homo si Pantog Pantog Pantog Homo si Homo si Nomasc	producing significant alignments         Download           41 sequences selected         Description           TED: Pan paniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         stelens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X, mRNA           TED: Pan tradictive endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X, mRNA         TED: Pan tradictive endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X, mRNA           TED: Pan tradictive endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X, mRNA         TED: Pan tradictive endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X, mRNA           construct DNA. Clone: pEN21AE 1261. Homo sapiens ERV3 open for endogenous retrovirus accestral sequence 3, without stop cod telens cDNA FL160255 complete cds, hightly similar to HERVR. 7a21.2 provirus ancestral Env polytotelin precursor podvies BAC done CH251-565C10 from chromosome 7, complete sequence           codvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           telens BAC clone RP11-460N20 from 7, complete sequence           odvies BAC clone RP11-460N20 from 7, complete sequence           plens cDNA FL123844 fis. clone LING13819. highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           us leucogenous retrovirus endogenous retrovirus endogenous retrovirus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA	Max Score 278 278 278 278 278 278 278 278 278 278	Gene Gene Cotal Score 278 278 278 278 278 278 278 278 278 278	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	Graphic           Graphic           E           value           7e-71	Show [ Dista Per. Ident 100.00% 100.0	1000 ✔ Accession Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AE5909991 AK2951891 AC1847992 AC1847992 AC1483133 NM 001135588 AC073210.8 AK074464.1 M12140.1 NM 001308194
Sequences Select all PREDIC Homoss PREDIC Synthetic Homoss Pantrog Pantrog Pantrog Homoss Homoss Homoss Subhon c Subhon c	producing significant alignments         Download           41 sequences selected         Description           FED: Pan paniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         pipens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           reconstruct DNA, clone, bRN21AE1261, Horno sapiens ERV3 open for endogenous retroviral sequence 3, without stop cod pipens cDNA FLJ60255 complete cds, hightly similar to HERVR, 7a212 provirus ancestral Env polyprotein precursor odytes BAC done CH251-565C10 from chromosome 7, complete sequence 0 odytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           toptes BAC done CH251-565C10 from chromosome 7, complete sequence 0 odytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           toptes BAC done CH251-565C10 from chromosome 7, complete sequence           odytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           toptens BAC clone RP11-4601020 from 7, complete sequence           uplens cDNA FLJ2384 fis, clone LN013819, highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           us leuccogenys endogenous retrovirus group 3	Max Score 278 278 278 278 278 278 278 278 278 278	GenE           GenE           Total           Score           278           272           272	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	Graphic           Graphic           E           value           7e-71	Show Show Per. Ident 100.00% 100	1000 ✔ 0 nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010956774 AE5909991 AC1483133 NM 001135588 AC1483133 NM 001135588 AC073210.8 AK074464.1 M12140.1 NM 001308194 AJ862653.1
Sequences Sequences Select all Homoss PREDIC PREDIC Swithels Homoss Pantrog Pantrog Homoss Ho	producing significant alignments         Download           41 sequences selected         Description           FED: Pan paniscus endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         pipens endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         FED: Pan troolodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           reconstruct DNA, clone, bRN21AE1261, Horno sapiens ERV3 open for endogenous retroviral sequence 3, without stop cod pipens cDNA FLJ60255 complete cds, hightly similar to HERVR, 7a212 provirus ancestral Env polyprotein precursor odytes BAC done CH251-565C10 from chromosome 7, complete sequence 0 odytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           replems cDNA FLJ60255 complete cds, bightly similar to HERVR, 7a212, provirus ancestral Env polyprotein precursor odytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           replems cDNA FLJ602610 from Ch:013819, hightly similar to Human endogenous retrovirus ERV3, pol-env-3LTR region           us leuccogenys endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           indogenous retrovirus proviral group 3 member 1, envelope (ERV3-1), mRNA           indogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA	Max Score 278 278 278 278 278 278 278 278 278 278	GenE           70tal           Score           278           272           272           272           272           272           272           272           272           272           272           273	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	Ims            Graphic         E           Value         7           7e-71         3           7e-71         3           7e-71         3           7e-71         3           7e-71         3           3e-69         3           3e-69         2	Show         E           15         Dista           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         90.33%           98.67%         98.67%	1000 ✔ 0 nce tree of resu Accession XM 034964253 NM 001007253 XM 024357564 XM 010955774 AC14357891 AC1483133 NM 001135588 AC1483133 NM 001135588 AC073210.8 AK074464.1 M12140.1 NM 001308194 AJ862653.1 XM 023196289
Sequencess Sequencess Select all Homoss PREDIC PREDIC PREDIC Southetic Homoss Pantog Pantog Pantog Homoss H	producing significant alignments         Download           41 sequences selected         Description           TED: Pan paniscus endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA         pipens endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12, mRNA           TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12, mRNA         TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12, mRNA           TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12, mRNA         TED: Pan troolodytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). transcript variant 12, mRNA           construct DNA, clone: pRN21AE1261. Homo sapiens ERV3 open for endogenous retroviral sequence 3, without stop cod pipens cDNA FLJ60255 complete cds. hightly similar to HERV-R. 7a212 provirus ancestral Env polyprotein precursor odytes BAC done CH251-656C10 from chromosome 7. complete sequence 0 odytes endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           tiplens CDNA FLJ60284 fis. Cone LNC13819. hightly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           us leucogenvs endogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           indogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           mdogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           mdogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           mdogenous retrovirus group 3 member 1. envelope (ERV3-1). mRNA           mdogenous r	Max Score 278 278 278 278 278 278 278 278 278 278	GenE           70tal           Score           278           272           272           267	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	Ims            Graphic         E           Value         7           7e-71         3           7e-71	Show         Istance           Istance         Distance           Ident         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         90.33%           98.67%         98.67%	1000 ✔ 0 nce tree of resul Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774 AE5909991 AC1483133 AC1483133 AC1483133 AC1483133 AC1483133 AC1483133 AC073210.8 AK074464.1 M12140.1 NM_001308194 AJ862653.1 XM_023196289 XM_031662056
Sequencess Sequencess Select all Homoss PREDIC PREDIC Synthetic Homoss Pantog Pantog Pantog Homoss Homoss Homoss Homoss Homoss Pantog P	producing significant alignments         Download           41 sequences selected         Description           TED: Pan paniscus endogenous retrovirus, group 3 member 1, envelope (ERV3-1), mRNA         pipens endogenous retrovirus, group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           TED: Pan tropiodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         TED: Pan tropiodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           TED: Pan tropiodytes endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         Construct DNA, clone, pRN21AE1261. Homo sapiens ERV3 gene for endogenous retroviral sequence 3, without stop cod pipens cDNA FLJ60255 complete cds, highly similar to HERV-R, 7a212 provirus ancestral Env polyprotein precursor odytes BAC done CH251-565C10 from chromosome 7, complete sequence addres endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           tiplens CDA FLJ60255, complete value 1, envelope (ERV3-1), mRNA         pipens CDAFLJ238415, clone LNG13819, highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           usiencogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         pipens CDAFLJ238415, clone LNG13819, highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           us leuccogenys endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         pipens cDAFLJ238415, clone LNG13819, highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region           us leuccogenys endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA         pipens cDAFLJ238415, clone	Max Score 278 278 278 278 278 278 278 278 278 278	Content of the second s	Column ank Query Cover 100% 100% 100% 100% 100% 100% 100% 100	E         E           Value         7e-71           7e-71         7e-71           7e-72         7e-71           7e-74         7e-71           7e-75         7e-71           7e-71         7e-72           7e-72         7e-74           7e-74         7e-74           7e-75         7e-74           7e-74         7e-74           7e-74	Show         Istance           Istance         Distance           Ident         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         99.33%           98.67%         98.67%	1000 ✔ 0 nce tree of resul Accession XM_034964253 NM_001007253 XM_024357564 XM_016956774 AE5909991 AC1483133 NM_001135588 AC0732108 AK074664.1 M12140.1 NM_001308194 AJ862653.1 XM_023196289 XM_031662056 XM_017956680
Sequences Sequences Select all PREDIC Homosi PREDIC Southeas Precipic Southeas Pantog	producing significant alignments         Download           41 sequences selected         Description           TED: Pan paniscus endogenous retrovirus, group 3 member 1, envelope (ERV3-1), mRNA         pipens endogenous retrovirus, group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           TED: Pan troolodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA         TED: Pan troolodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant 12, mRNA           TED: Pan troolodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X2, mRNA         TED: Pan troolodvies endogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X1, mRNA           construct DNA, clone, pRN21AE1281. Homo sapiens ERV3 gene for endogenous retrovirus accestral Env polyprotein precursor odytes BAC done CH251-632C19 from chromosome 7, complete sequence         avita secuence           odvies BAC done CH251-632C10 from chromosome 7, complete sequence         advies endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           pipens CDNA FL/3884 fis, clone LNG13819. highly similar to Human endogenous retrovirus ERV3, pol-env-31TR region         us leucogenvis endogenous retrovirus group 3 member 1, envelope (ERV3-1), mRNA           midgenous retrovirus proviral envelope ordein         TED: Pillocolobus tertovirus group 3 member 1, envelope (ERV3-1), mRNA           modgenous retrovirus proviral group 3 member 1, envelope (ERV3-1), transcript variant X1, mRNA         mdogenous retrovirus group 3 member 1, envelope (ERV3-1), transcript variant X1, mRNA	Max Score 278 278 278 278 278 278 278 278 278 278	Content of the second s	Colum ank Query Cover 100%	Staphic           Staphic           E           value           7e-71           3e-69           3e-69           2e-67           2e-67           2e-67           2e-67           2e-67	Show         Istance           Istance         Distance           Ident         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         100.00%           100.00%         99.33%           98.67%         98.67%           98.67%         98.67%	1000 ✔ 0 nce tree of resu Accession XM. 034964253 NM. 001007253 XM. 024357564 XM. 024357564 XM. 010956774 AE590999.1 AK295189.1 AC148313.3 NM. 001135588 AC073210.8 AK074464.1 M12140.1 NM. 001308194 AJ852653.1 XM. 023196289 XM. 031662056 XM. 017956680 XM. 030932630

Fig.4b A BLAST analysis on reads sampled from the 100% hit results confirmed that it was found only in humans. Once again confirming human origin.

Description	gnl SRA SRR11119759.706		to		to				to			
Molecule type	dna								_			
Query Length	150									Fi	ilter	Reset
Other reports	Distance tree of results											
Descriptions	Graphic Summary	Alignments	Taxonomy									
Sequences p	producing significant a	lignments			Down	load 🗸 🛛	Manage	Colu	mns ~	Sho	w 100	0 🗸 🛛
select all	1 sequences selected						Gent	<u>Bank</u>	<u>Graph</u>	<u>ics D</u>	istance ti	ree of results
		De	escription				Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Synthetic c	onstruct DNA, clone: pFN21AE12	261, Homo sapiens E	RV3 gene for endog	enous retroviral sequer	nce <u>3. withou</u>	t stop codon, in	278	278	100%	7e-71	100.00%	AB590999.1

#### Fig.4c The sequence have no matches outside of Primates.

select all 100 sequences selected Graphics									
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession	
	SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168375.5045789.1	
	SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168375.5964.1	
Des	scription	Homo sapiens BAC clone RP11-460N20 from 7, complete $\operatorname{seq}$							
Мо	lecule type	nucleic acid							
Qu	ery Length	203396							
Oth	er reports	Distance tree of results MSA viewer 😢							

#### Fig.5a <u>SRX6893156</u> also returned 100% matched results from the human Genome.

<ul><li>✓</li></ul>	select all 14 s	equences selected	Gen	<u>Bank</u>	<u>Grap</u>	<u>hics</u>	Distance	tree of results
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Homo sapiens	3AC clone RP11-460N20 from 7, complete sequence	278	278	100%	6e-71	100.00%	AC073210.8
	Pan troglodytes	BAC clone CH251-623C19 from chromosome 7, complete sequence	267	267	100%	1e-67	98.67%	AC184799.2
	Pan troglodytes	BAC clone CH251-2015 from chromosome 7, complete sequence	267	267	100%	1e-67	98.67%	AC174000.3
≤	Pan troglodytes	BAC clone CH251-565C10 from chromosome 7, complete sequence	267	267	100%	1e-67	98.67%	AC148313.3
Desc	ription	gnl SRA SRR10168375.5045789.1 5045789 (Biological)						
Mole	ecule type	dna						
Quei	ry Length	150						
Othe	er reports	Distance tree of results MSA viewer 😧						
Des	scription	gnl SRA SRR10168375.5964.1 5964 (Biological) to		to			to	
Мо	lecule type	dna	]					
Qu	ery Length	150					Filter	Reset
Oth	her reports	Distance tree of results 😨						
	Descriptions	Graphic Summary Alignments Taxonomy						
:	Sequences	producing significant alignments Download $^{\vee}$	Mana	ge Co	lumns	~ <u>s</u>	how 1	1000 🗸 🕜
	select all	1000 sequences selected	G	enBanl	<u>Gra</u>	<u>phics</u>	Distanc	e tree of results
		Description M Sc	ax To ore Sco	tal Qu ore Co	iery iver <b>va</b>	E lue	Per. Ident	Accession
	Homo sap	iens IncAB572.1 IncRNA gene, complete sequence 21	78 27	8 10	0% 7e	-71 10	00.00% <u>N</u>	IK280613.1
	Pan troglo	dytes chromosome 2 clone CH251-60P06, complete sequence 21	78 27	78 10	0% 7e	-71 10	0.00% <u>A</u>	C279084.1
	Pan troglo	dytes chromosome 2 clone CH251-17O22, complete sequence 21	78 27	78 10	0% 7e	9-71 10	00.00% <u>A</u>	C278930.1
	Pan troglo	dytes chromosome 2 clone CH251-108A24, complete sequence 21	78 27	78 10	0% 7e	-71 10	00.00% <u>A</u>	C278921.1
	Eukaryotic	synthetic construct chromosome 21 21	78 79	38 10	0% 7€	-71 10	0.00% C	P034500.1

Fig.5b BLAST search on the result returned 100% match only found in humans. Confirming origininhuman-derivedmaterial.

Description	gnl SRA SRR10168375.5964	4.1 5964 (Biologica	l)	to			to			to	
Molecule type	dna										
Query Length	150									Filter	Reset
Other reports	Distance tree of results										
Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences p	producing significant al	lignments		Down	load ~	Mana	ge Colu	mns	∽ Sh	ow 100	0 🗸 🕜
select all	14 sequences selected					Ge	n <u>Bank</u>	<u>Grap</u>	hics .	Distance t	ree of results
		De	scription			Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Eukaryotic	synthetic construct chromosome	21				278	7938	100%	7e-71	100.00%	CP034500.1
Eukaryotic	synthetic construct chromosome	13				267	14570	100%	1e-67	98.67%	CP034516.1
Eukaryotic	synthetic construct chromosome	16				267	10333	100%	1e-67	98.67%	CP034494.1
Eukaryotic	synthetic construct chromosome	15				267	9021	100%	1e-67	98.67%	CP034493.1
Eukaryotic	synthetic construct chromosome	13				267	14570	100%	1e-67	98.67%	CP034491.1
Eukaryotic	synthetic construct chromosome	18				261	15047	100%	7e-66	98.00%	CP034496.1
Eukaryotic	synthetic construct chromosome	17				261	6545	100%	7e-66	98.00%	CP034495.1
Eukaryotic	synthetic construct chromosome	20				219	7949	98%	4e-53	93.79%	CP034499.1
Eukaryotic	synthetic construct chromosome	19				209	3521	96%	3e-50	93.10%	CP034522.1
Eukaryotic	synthetic construct chromosome	19				209	3766	96%	3e-50	93.10%	CP034497.1
Eukaryotic	synthetic construct chromosome	22				207	2291	96%	9e-50	92.47%	CP034501.1
Eukaryotic	synthetic construct chromosome	14				207	13851	96%	9e-50	92.47%	CP034492.1
Gossypiun	n hirsutum clone NBRI_GE27093	3 microsatellite seque	ince			189	189	96%	3e-44	90.41%	JX591845.1
Eukaryotic	synthetic construct chromosome	Y				180	2399	96%	2e-41	89.44%	CP034510.1

Fig.5c BLAST result of the sequences in question revealed that it is not found outside of Primates.

C	se	lect all	100 sequences selected					Graph	nics Distance tree of results
			Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
5		SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168376.17339580.1
5		SRX6893155		278	278	0%	2e-69	100.00%	SRA:SRR10168376.17013625.2
5		SRX6893155		278	278	0%	2e-69	100.00%	SRA:SRR10168376.17013625.1
5	4	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.16930714.2
5	2 5	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.16930714.1
5		SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.15267479.2
5		SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168376.15267479.1
5		SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168376.13985702.2
2		SRX6893155		278	278	0%	2e-69	100.00%	SRA:SRR10168376.13985702.1
5		SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.13353823.2
5		SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.13353823.1
5	2 3	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.11109740.1
2		SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.9343845.2
5		SRX6893155		278	278	0%	2e-69	100.00%	SRA:SRR10168376.9232549.2
Desc	cript	ion	Homo sapiens BAC clone RP11-460N20 from 7, complete seq						
Mole	ecule	e type	nucleic acid						
Que	ry Le	ength	203396						
Othe	er re	ports	Distance tree of results MSA viewer 😮						

Fig.6a Similarly, BLAST research on <u>SRX6893155</u> gives large number of full length 100% matches to the human genome.

$\checkmark$	select all 57 se	equences selected					GenBan	<u>k</u> <u>G</u>	<u>raphics</u>	<u>Distan</u>	ice tree of results
			Descri	ption		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Homo sapiens F	OSMID clone ABC13-48840700E	15 from chromosome	7, complete sequence		278	278	100%	6e-71	100.00%	AC242196.4
	Pan troglodytes	BAC clone CH251-340l24 from ch	iromosome 7 <u>, comp</u> le	ete sequence		278	278	100%	6e-71	100.00%	AC185242.2
	Pan troglodytes	BAC clone CH251-623C19 from c	hromosome 7 <u>, compl</u>	lete sequence		278	278	100%	6e-71	100.00%	AC184799.2
≤	Pan troglodytes	BAC clone CH251-114G16 from c	hromosome 7, compl	lete sequence		278	278	100%	6e-71	100.00%	AC183835.2
	Pan troglodytes	BAC clone CH251-2O15 from chr	omosome 7, complete	e sequence		278	278	100%	6e-71	100.00%	AC174000.3
	Homo sapiens B	AC clone RP11-47909 from 7, co	mplete sequence			278	278	100%	6e-71	100.00%	AC073107.7
	Pan troglodytes	BAC clone CH251-565C10 from c	hromosome 7, compl	lete sequence		278	278	100%	6e-71	100.00%	AC148313.3
<b>_</b>	Homo sapiens B	AC clone RP11-460N20 from 7, c	omplete sequence			278	278	100%	6e-71	100.00%	AC073210.8
<b>~</b>	PREDICTED: Ce	ebus capucinus imitator small inte	g <u>ral membrane protei</u>	in 11A (SMIM11A), transc	cript variant X6, mRNA	87.9	87.9	49%	1e-13	88.00%	<u>XM_017526193.1</u>
De	escription	gnl SRA SRR10168376.15	267479.2 1526747	79 (Biological)							
M	olecule type	dna									
Q	ery Length	150									
01	her reports	Distance tree of results	MSA viewer 🔞								
De	scription	gnl SRA SRR10168376.1398	5702.1 13985702 (	(Biological)	to		to			+	
Мо	lecule type	dna		()			10			`	0
Ou	erv Length	150			-					Filter	Reset
	,										
Ot	ier reports	Distance tree of results 🔞									
Ot	ter reports	Distance tree of results									
Oti	ner reports Descriptions	<b>Distance tree of results O</b> Graphic Summary	Alignments	Taxonomy							
Otl	ner reports Descriptions Sequences p	Distance tree of results Graphic Summary roducing significant al	Alignments ignments	Taxonomy	Download 🗸	Ма	nage Co	olumn	s ~ :	Show	1000 🗸 😢
Ot	Descriptions Sequences p ✓ select all &	Distance tree of results Graphic Summary roducing significant all 87 sequences selected	Alignments ignments	Taxonomy	Download 🗸	Ма	nage Co	olumn <u>ık G</u>	s Y s	Show	1000 V 2
Ot	Descriptions Sequences p Select all s	Distance tree of results Graphic Summary roducing significant all 97 sequences selected	Alignments ignments	Taxonomy	Download 🗠	Ma Max Score	nage Co GenBar Total C Score C	olumn <u>ak G</u> Query Cover <b>X</b>	s × s raphics E ralue	Show Distant Per.	1000 V 2 ce tree of results Accession
Oti	Pescriptions Descriptions Sequences p Select all & Homo sable	Distance tree of results Graphic Summary roducing significant all 87 sequences selected	Alignments ignments Descr	Taxonomy	Download 🗸	Max Score 278	nage Co GenBar Total C Score C 278 1	olumn <u>ik G</u> Duery Cover N 00%	s × s raphics E raiue 7e-71 1	Distant Per. Ident	1000 V 2 ce tree of results Accession AC073210.8
Ot	Pescriptions Descriptions Sequences p Select all a Homo sabis Pan troolod	Distance tree of results Graphic Summary roducing significant all 87 sequences selected ens BAC clone RP11-460N20 fro Mes BAC clone CH251-623C191	Alignments ignments Descr m 7. complete seque from chromosome 7.	Taxonomy ription ence complete sequence	Download 🗸	Max Score 278 272	nage Co GenBar Total C Score C 278 1 272 1	blumn <u>Ik G</u> Duery Cover V 00%	s ∨ S raphics E value 7e-71 1 3e-69 S	Show Distant Per. Ident 00.00% ( 99.33% (	1000 ✓ ♀ ce tree of results Accession AC073210.8 AC184799.2
Oti	Terreports         Descriptions         Sequences p         Select all s         Homo saple         Pan troolod         Pan troolod         Pan troolod	Distance tree of results Graphic Summary roducing significant all 97 sequences selected ens BAC clone RP11-460N20 fro Mes BAC clone CH251-623C191 Mes BAC clone CH251-565C101	Alignments ignments Descr m 7. complete seque form chromosome 7. tom chromosome 7.	Taxonomy ription ance complete sequence complete sequence	Download 🗸	Max Score 278 272 272	nage Co GenBar Total C Score C 278 1 272 1 272 1	olumn ak <u>G</u> auery cover v 00% 3 00% 3	s × 3 raphics E ralue 7e-71 1 3e-69 3 3e-69 3	Distant           Per.         Ident           00.00%         2           99.33%         2	1000 ✔ 2 ce tree of results Accession AC073210.8 AC184799.2 AC148313.3
Oti	erreports  Descriptions  Sequences p  Select all s  Homo sapis  Pan troolod  Macaca mul	Distance tree of results Graphic Summary roducing significant all 97 sequences selected ens BAC clone RP11-460N20 fro Mes BAC clone CH251-623C191 Mes BAC clone CH251-206B6 (childre	Alignments ignments Descr m7.complete seque from chromosome 7. from chromosome 7.	Taxonomy ription ance complete sequence complete sequence d Research Institute Rhe	Download ~	Max Score 278 272 272 272 217	nage Co GenBar Total C Score C 278 1 272 1 272 1 272 1 272 1	olumn ak G over V 00% 3 00% 3 00% 3	s × 3 raphics E ralue 7e-71 1 3e-69 9 3e-69 9 2e-52 9	Distant           Per.           Ident           00.00%           99.33%           99.33%           92.67%	1000 ✔ 2 ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6
Oth	erreports  Descriptions  Sequences p  Select all s  Pan troolod  Pan troolod Macaca mul Homo saple Homo saple	Distance tree of results Graphic Summary roducing significant al 97 sequences selected ens BAC clone CH251-823C191 Mes BAC clone CH251-823C191 Mes BAC clone CH251-206B6 (childre ens tripartite motif containing 24.1	Alignments ignments Descr m7.complete seque from chromosome 7. ms.Hospital Oakland (TRIM24).RefSeqGer	Taxonomy ription ance complete sequence complete sequence J Research Institute Rhe ne on chromosome 7	Download ~	Max Score 278 272 272 217 182	GenBar           Total         C           Score         C           278         1           272         1           272         1           217         1           182         9	blumn           ak         G           buery         v           cover         v           00%         v	s ∨ 5 raphics F ralue 7e-71 1 3e-69 5 3e-69 5 2e-52 5 5e-42 5	Distant           Per.           Ident           09.33%           99.33%           92.67%           88.59%	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6 NG_023286.1
Ot	erreports  Descriptions  Sequences p  Select all s  Pan troolod  Pan troolod  Macaca mul Homo sapli Homo sapli Homo sapli	Distance tree of results Graphic Summary roducing significant al 97 sequences selected ens BAC clone CH251-823C191 Mes BAC clone CH251-823C191 Mes BAC clone CH251-206B0 (childre ens tripartite motif containing 24) ens chromosome 7 clone RP11-	Alignments ignments Descr m7. complete seque from chromosome 7. m5. Hospital Oaklanc (TRIII/24). RefSeaGer 199L 18. complete se	Taxonomy ription ance complete sequence complete sequence I Research institute Rhe ne on chromosome 7 isquence	Download ~	Max Score 278 272 272 217 182 182	nage Co GenBar Total C 278 1 272 1 272 1 217 1 182 9 182 9	blumn           k         G           cover         V           000%         Cover           000% <th>s &gt; 3 raphics F ralue 7e-71 1 3e-69 9 3e-69 9 2e-52 9 5e-42 9 5e-42 9</th> <th>Distant           Per.           Ident           00.00%           99.33%           92.67%           88.59%</th> <th>1000 ✔ 2 Ce tree of results Accession AC073210.8 AC183133 AC210125.6 NG_023286.1 AC013429.12</th>	s > 3 raphics F ralue 7e-71 1 3e-69 9 3e-69 9 2e-52 9 5e-42 9 5e-42 9	Distant           Per.           Ident           00.00%           99.33%           92.67%           88.59%	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC183133 AC210125.6 NG_023286.1 AC013429.12
Oth	enerreports  Descriptions  Sequences p  Select all  Enerreport En	Distance tree of results Graphic Summary roducing significant al 97 sequences selected ens BAC clone CH251-632C191 Mes BAC clone CH251-565C101 liatta BAC CH250-206B6 (childre ens tripartite motif containing 241 ens chromosome 7 clone RP11- ens chromosome 7 clone RP11-	Alignments ignments Descr m7. complete seque from chromosome 7. m5. Hospital Oakland (TRIII/24). RefSeaGer 199L 18. complete se 256C24. complete se	Taxonomy ription ance complete sequence complete sequence d Research institute Rhe ne on chromosome 7 requence aquence	Download ~	Max Max Score 278 272 272 217 182 182 182	GenBar           Total         C           Score         C           2778         1           2772         1           2772         1           2172         1           2172         1           2182         1           182         1           182         1	Ik         G           Nuery         N           Nuery         N           000%         1           000%         1           000%         2	s × 5 raphics Falue 7e-71 1 3e-69 5 3e-69 5 2e-52 5 5e-42 5e-56 5e-56 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-566 5e-56	Show         Distant           Per.         Ident           Ident         99.33%           99.33%         4           99.33%         4           98.59%         4           88.59%         4           88.59%         4	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC18313.3 AC210125.6 NG_023286.1 AC013429.12 AC008265.15
Oth		Distance tree of results Graphic Summary roducing significant al 97 sequences selected ens BAC clone CH251-623C191 Mes BAC clone CH251-623C191 Mes BAC clone CH251-565C101 liatta BAC CH250-206B6 (Childre ens tripartite motif containing 24.1 ens chromosome 7 clone RP11- ens chromosome 7 clone RP11- synthetic construct chromosome	Alignments ignments Descr m7.complete seque from chromosome 7. m5.Hospital Oakland (TRIII/24).RefSeaGer 199L18.complete se 256C24.complete ser Y	Taxonomy ription ance complete sequence complete sequence I Research institute Rhe ne on chromosome 7 requence aquence	Download ~	Max Score 278 272 272 217 182 182 182 182 176	GenBar           Total         C           Score         C           2778         1           272         1           272         1           272         1           182         1           182         1           182         3	blumn           k         G           tuery         V           cover         V           000%         C           00	<b>s</b> ∨ 2 <b>raphics</b> <b>F</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphics</b> <b>raphic</b>	Distant           Per.         Ident           Ident         99.33%           99.33%         98.59%           88.59%         98.59%           88.59%         98.859%	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC18313.3 AC210125.6 NG_023286.1 AC013429.12 AC008265.15 CP034510.1
Ott	Pescriptions  Sequences p  Select all s  Pan troglod Pan troglod Pan troglod Macaca mul Homo sapli Homo sapli Lucaspid Lucaspid Eukanyolics Select all 3 seq	Distance tree of results Graphic Summary roducing significant all 87 sequences selected ens BAC clone RP11-460N20 fro httes BAC clone CH251-623C19 f httes BAC clone CH251-632C19 f httes	Alignments ignments Descr m 7. complete seque from chromosome 7. in's Hospital Oaklano (TRIM24). RefSeaGer 1991 18. complete se 256C24. complete set Y	Taxonomy ription ence complete sequence complete sequence d Research Institute Rhe ne on chromosome 7 isquence aquence	Download ~	Max Score 278 272 272 217 182 182 182 176	GenBar           Total         C           Score         C           278         1           272         1           272         1           272         1           182         1           182         1           353         1           GenBar         GenBar	Ik         G           Nuery         V           Sover         V           000%         T	s         S	Show         Distant           Per.         Ident           100.00%         4           99.33%         4           92.67%         4           88.59%         4           88.59%         4           87.92%         9	1000 ✓ 2 ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6 NG.023286.1 AC013429.12 AC008265.15 CP034510.1 ce tree of results Accession
oti	Pescriptions Sequences p Sequences p Select all s Pan troolod Pan troolod Pan troolod Macaca mul Homo sapis Homo sapis Homo sapis Eukaryotic s Select all 3 seq	Distance tree of results Graphic Summary roducing significant all 87 sequences selected ens BAC clone CH251-623C19 I Wes BAC clone CH251-623C19 I Wes BAC clone CH251-623C19 I Wes BAC clone CH251-623C19 I when BAC character and the selection of the selection is atta BAC CH250-206B6 (Childre ens tripartite motif containing 24.1 ens chromosome 7 clone RP11- ens chromosome 7 clone RP11- ent ch	Alignments ignments Descr m7.comolete seque tom chromosome 7. tom	Taxonomy	Download ~	Max Score 278 272 277 217 182 182 182 182 176	GenBar           Total         C           Score         C           278         1           272         1           272         1           272         1           272         1           272         1           182         9           353         9           GenBar         GenBar           tax         Tc           Score         Score	Image: style	s ∨ 2 raphics E Falue 1 3e-69 2 3e-70 2 3e 3e-70 2 3e-70 2 3e-70 2 3e-70 2 3e-70 2 3e-70 2 3e	Distant           Per.           Ident           00.00%           99.33%           88.59%           88.59%           97.792%           Distant	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6 NG_023286.1 AC013429.12 AC008265.15 CP034510.1 AC008265.15 CP034510.1 Accession nt Accession
	Pescriptions  Sequences p  Select all s  Pan troolod Pan troolod Pan troolod Macaca mul Homo saple Homo saple Homo saple Lucaryotic s  Select all 3 seq Eukaryotic synth	Distance tree of results Graphic Summary roducing significant all 87 sequences selected ens BAC clone CH251-623C19 I Mes Chromosome 7 clone RP11- ens chromosome	Alignments ignments Descr m7.complete seque form chromosome 7. tom chromosome 7. to	Taxonomy ription ence complete sequence complete sequence d Research institute Rhe ne on chromosome 7 equence equence	Download ~	Max Score 278 2772 2772 2177 182 182 182 182 176	GenBar       Total     C       Score     C       278     1       272     1       272     1       272     1       182     9       182     9       353     9       GenBar     Score       tax     Tc       Score     Score       176     3	blumn           tk         G           cover         V           cover         C           cover         C           cover         C	s ✓ 2 raphics F ralue 7e-71 1 3e-69 9 2e-52 9 2e-	Distant           Per           Ident           00.00%           99.33%           99.33%           98.59%           98.59%           Distant           Distant           Distant           Distant           Ident	1000 ✔ 2 Ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6 NG_023286.1 AC013429.12 AC008265.15 CP034510.1 Accession nt Accession 2% CP034510.1
	Pescriptions  Sequences p  Sequences p  Sequences p  Pantroolod Pantroolod Pantroolod Macaca mu Homo sapis Homo sapis Homo sapis Eukaryotic synth Eukaryotic synth	Distance tree of results Graphic Summary roducing significant all 87 sequences selected ens BAC clone CH251-623C19 I Wes BAC clone CH251-625C10 I latta BAC CH250-206B6 (Childre ens tripartite motif containing 24 I ens chromosome 7 clone RP11- ens	Alignments ignments Descr m 7, complete seque from chromosome 7. trom chromosome	Taxonomy riplion ance complete sequence complete sequence d Research institute Rhe ne on chromosome 7 requence equence	Download ~	Max Score 278 272 272 277 182 182 182 182 176	GenBar           Total         C           Score         C           278         1           272         1           272         1           272         1           272         1           182         9           353         9           GenBar         T           tax         TC           Score         Score           176         3           143         1	blumn           ik         G           cover         V           cover         C	s            raphics            raphics            rau            7e-71         1           3e-69            3e-69            2e-52            3e-69            3e-69            age-40            age-40         <	Distant           Per           Ident           00.00%           99.33%           92.67%           98.59%           98.859%           0           0           0           0           10           0           10           10           11           12           12           13           14           14           14           14           14           14           14	1000 ✔ 2 ce tree of results Accession AC073210.8 AC184799.2 AC148313.3 AC210125.6 NG_023286.1 AC013429.12 AC008265.15 CP034510.1 Accession rt. Accession 2% CP034510.1 0% CP034492.1

Fig.6b The results, when put through BLAST, confirms that the 100% matches are in fact derived from a Hominid origin.

Description Molecule type Query Length Other reports	Homo sapiens BAC clone RP11-460N20 from 7, complete seq       Kolecule type     nucleic acid       203396       Distance tree of results     MSA viewer					Eva	lue	to		Query Co	to	Reset
Descriptions Sequences p	Graphic Summary	Alignments		Dow	nload	~	Mar	nage Co	olumns	✓ Show	100	• 0
🗹 select all	100 sequences selected								Graph	nics <u>Dista</u>	nce tree c	of results
		Description	I		Max Score	Total Score	Query Cover	E value	Per. Ident	,	Accession	
SRX689315	<u>3</u>				278	278	0%	2e-69	100.00%	SRA:SRR1	168378.18	32954.1

Fig.7a <u>SRX6893153</u> have also returned 100% match full-length read on this tiny part of the human genome.

Description	gnl SRA SRR10168378.18	Percent Identity E	value			Query	Covera	ige			
Molecule type	dna			to	to				to	_	
Query Length	150										
Other reports	Distance tree of results	MSA viewer 🔞						F	ilter	Rese	<b>۱</b>
Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences	producing significant a	alignments		Download 🗡	Manage	Colun	nns ~	Sho	w 10	00 🗸	0
🗹 select all	170 sequences selected				Gen	<u>Bank</u>	Graph	ics D	istance	tree of res	ults
		De	scription		Max Score	Total Score	Query Cover	E value	Per. Ident	Access	ion
Homo sapi	ens FOSMID clone ABC18-86211	1 from chromosome 7, co	omplete sequence		278	278	100%	6e-71	100.00%	AC24520	<u>5.1</u>
Homo sapi	ens FOSMID clone ABC13-48840	700E15 from chromoson	ne 7 <u>, complete sequence</u>		278	278	100%	6e-71	100.00%	AC24219	5.4
Homo sapi	ens BAC clone RP11-460N20 from	n 7. complete sequence			278	278	100%	6e-71	100.00%	AC07321	<u>).8</u>
Pan troglo	dytes BAC clone CH251-487D11 f	rom chromosome 7, con	nplete sequence		272	272	100%	3e-69	99.33%	AC18273	3.3
select all	7 sequences selected					Gen	<u>Bank</u>	Grap	hics [	Distance t	ree of re
		C	Description			Max Score	Total Score	Query Cover	E value	Per. Ident	Access
Eukaryotic s	ynthetic construct chromoson	ne 15				211	211	98%	7e-51	92.11%	CP0344
Eukaryotic s	synthetic construct chromoson	ne 16				206	1214	98%	3e-49	91.45%	CP0344
Eukaryotic s	ynthetic construct chromoson	ne 13				200	200	98%	2e-47	90.79%	CP0345
Eukaryotic s	ynthetic construct chromoson	ne 21				200	401	98%	2e-47	90.79%	CP0345
Eukaryotic s	synthetic construct chromoson	ne 13				200	200	98%	2e-47	90.79%	CP0344
Eukaryotic s	synthetic construct chromoson	ne 18				195	195	96%	7e-46	90.60%	CP0344
Eukarvotic s	withetic construct chromoson	ne 17				195	195	98%	7e-46	90,13%	CP0344

Fig.7b Similarly, the read is only found in humans—indicating the Homo Sapiens Trace result is

accurate.

escription	Homo sapiens BAC clone RP11-450O3 from 7, complete sequence						
Iolecule type	nucleic acid						
uery Length	195834						
)ther reports	Distance tree of results MSA viewer 😵						
Descriptions	Graphic Summary Alignments						
Sequences p	producing significant alignments	Dot	wnload	~	Manage	Column	s ~ Show 100 🗸
select all	100 sequences selected					<u>Gra</u>	phics Distance tree of res
select all	100 sequences selected Description	I	Max To Score So	otal Que	y E r value	<u>Gra</u> Per. Ident	phics Distance tree of res
select all	100 sequences selected Description 54	l s	Max To Score Sc 278 5	otal Que core Cove 56 0%	y E r value 1e-69	Gra Per. Ident 100.00%	phics Distance tree of res Accession SRA-SRR10168377.163022
<ul> <li>select all</li> <li>SRX68931</li> <li>SRX68931</li> </ul>	100 sequences selected Description 54 54		Max To Score Sc 278 5 278 5	otal Que core Cove 56 0%	y E r value 1e-69 1e-69	<u>Gra</u> Per. Ident 100.00%	phics Distance tree of re: Accession SRA:SRR10168377.163022 SRA:SRR10168377.120148

Fig.8a Reads from the Human PMS1 gene is recovered from <u>SRX6893154</u> with a query sequence only 195834bp in length.

Description	gnl SRA SRR10168377.163	02266.1 16302266 (	Biological)						
Molecule type	dna							Filte	r Reset
Query Length	150								
Other reports	Distance tree of results								
Descriptions	Graphic Summary	Alignments	Taxonomy						
Sequences	producing significant a	lignments		Download 🗡	Mar	nage Col	umns 🗸	Show	1000 🗸 😨
select all	311 sequences selected					GenBank	<u>Graphi</u>	<u>cs Dista</u>	ance tree of results
		Descr	iption		Max 1 Score S	Total Qui	ery E ver value	Per. Ident	Accession
Homo sa	piens PMS1 homolog 2, mismato	h repair system comp	onent pseudogene 8	(PMS2P8) on chromosome 7	278	278 10	0% 7e-71	100.00%	NG_006447.3
Homo sa	piens PMS1 homolog 2, mismate	h repair system comp	onent pseudogene 1	) (PMS2P10) on chromosome 7	278	278 10	0% 7e-71	100.00%	NG_023454.4
Homo sa	piens PMS1 homolog 2, mismate	h repair system comp	onent pseudogene 6	(PMS2P6) on chromosome 7	278	278 10	0% 7e-71	100.00%	NG_006449.3
Homo sa	piens BAC clone CH17-264B6 fro	m chromosome 7, cor	mplete sequence		278	1112 10	0% 7e-71	100.00%	AC211476.5
Homo sa	piens BAC clone CH17-220H16 f	rom chromosome 7, c	omplete sequence		278	1112 10	0% 7e-71	100.00%	AC211491.5
Homo sa	piens FOSMID clone ABC10-455	5000F15 from chrome	osome 7, complete s	equence	278	556 100	0% 7e-71	100.00%	AC244146.2
<ul> <li>Homo sap</li> <li>Homo sap</li> <li>Homo sap</li> <li>Homo sap</li> </ul>	viens BAC done RP11-313P13 f viens chromosome 7 done VMR viens chromosome 7 done VMR viens chromosome 7 done CH1	rom 7, complete sequ C53-89F05, complete C62-404M06, complet 7-61K4, complete se	ience e sequence tte sequence quence		278 273 273 273	3 556 2 272 2 272 2 272 2 272	100% 7 100% 3 100% 3 100% 3	e-71 100 e-69 99 e-69 99 e-69 99	AC005488.2           .33%         AC278394.1           .33%         AC278331.1           .33%         AC270699.1
Homo sapi	iens PMS8 mRNA (yeast misma	ch repair gene PMS1	homologue), partial	cds (C-terminal region)	248	248 1	100% 6e-6	62 96.67	% <u>D38503.1</u>
Homo sapi	iens PMS1 homolog 2. mismatcl	n repair system comp	onent (PMS2), RefSe	gGene (LRG_161) on chromosome 7	244	244 1	100% 7e-6	61 96.00	% <u>NG 008466.1</u>
Canis lupu:	s familiaris breed Labrador retrie	ver chromosome 06a			154	154 8	4% 1e-3	3 88.89%	CP050586.1
Canis lupus	s familiaris breed Labrador retrie	ver chromosome 06b			154	154 8	4% 1e-3	3 88.89%	CP050622.1
Pipistrellus	pipistrellus genome assembly,	chromosome: 5			139	139 8	16% 3e-2	9 86.15%	LR862361.1
Synthetic co	onstruct Homo sapiens clone ccs	bBroadEn_14772 PM	<u>S2 gene, encodes co</u>	mplete protein	126	126 4	19% 3e-2	5 97.30%	KJ905275.1
	D: Zalophus californianus PMS1	homolog 2, mismatch	repair system comp	onent (PMS2), transcript variant X3, mRNA	124	124 5	i4% 9e-2	5 93.90%	XM 027610028.2
	D: Desmodus rotundus PMS1 ho	molog 2, mismatch re	pair system compor	ent (PMS2), transcript variant X2, mRNA	121	121 5	i2% 1e-2	3 94.87%	XM 024576931.1
	D: Desmodus rotundus PMS1 ho	molog 2, mismatch re	epair system compor	ent (PMS2), transcript variant X1, mRNA	121	121 5	i2% 1e-2	3 94.87%	XM 024576930.1
	D: Myotis lucifugus PMS1 homolo	og 2. mismatch repair	<u>system component (</u>	PMS2), transcript variant X4, mRNA	121	121 5	i2% 1e-2	3 94.87%	XM 023761457.1
PREDICTE	D: Myotis lucifugus PMS1 homolo	og 2, mismatch repair	system component (	PMS2), transcript variant X3, mRNA	121	121 5	i2% 1e-2	3 94.87%	XM 023761456.1

Fig.8b This PMS1 read is only found in Humans. This is clearly a contaminant from a hominid origin.

escription	Homo sapiens BAC clone	RP11-611L7 from 7, complete sequence	Percent Identity		E value			Query Coverage
lolecule type	nucleic acid		to			to		to
uery Length	173967							
ther reports	Distance tree of results	MSA viewer 😮						Filter Reset
Descriptions	Graphic Summary	Alignments						
Sequences	producing significant	alignments	Dowr	nload	× N	lanage C	olumns	∽ Show 100 ✔
🗹 select all	100 sequences selected						Grap	hics Distance tree of resul
		Description		Max T Score S	otal Que	ry E er value	Per. Ident	Accession
SRX689313	<u>39</u>			278	78 09	6 3e-69	100.00%	SRA:SRR10168392.39544030.
SRX689313	<u>39</u>			278	78 09	6 3e-69	100.00%	SRA:SRR10168392.28917809
SRX689313	<u>39</u>			278	78 09	6 3e-69	100.00%	SRA:SRR10168392.14357888
SRX689313	39			278	78 09	6 3e-69	100.00%	SRA:SRR10168392.2548655.7

Fig.9a Similarly, multiple 100% match Full length reads were obtained from <u>SRX6893139</u>. As this query sequence is only 173967 nucleotides in length, the real extent of Human-derived contamination is also extremely severe.

Description Molecule type Query Length Other reports	gnl SRA SRR10168392.2 dna 150 Distance tree of results	Percent Identity	E value	to		Qı	Jery Cov	to Reset		
Descriptions	Graphic Summary	Alignments	Taxonomy	Download Y	Маг	nage Co	olumn	s × s	Show	1000 🗸 🕄
select all	66 sequences selected	augiments			Max	GenBar Total	<u>nk G</u>	raphics F	Distan	ice tree of results
Homo sapi	iens zinc finger protein 316 (ZNF	Desc 316). mRNA	cription		Score 278	Score 278	Cover 100%	value 6e-71	Ident 100.00%	Accession <u>NM_001278559.2</u>
	ED: Homo sapiens zinc finger pro ED: Homo sapiens zinc finger pro	otein 316 (ZNF316), transc otein 316 (ZNF316), transc	<u>ript variant X3, mRNA</u> ript variant X2, mRNA		278 278	278 278	100% 100%	6e-71 6e-71	100.00% 100.00%	<u>XM_024446619.1</u> <u>XM_024446618.1</u>
PREDICTE     Homo sapi     PREDICTE     PREDICTE	ED: Homo sapiens zinc finger pro iens BAC clone RP11-611L7 fror ED: Pongo abelii zinc finger prote	otein <u>316 (ZNF316), transc</u> n <u>7. complete sequence</u> ein 316 (ZNF316), mRNA	<u>ript variant X1, mRNA</u>		278 278 272	278 278 272	100% 100% 100%	6e-71 6e-71 3e-69	100.00% 100.00% 99.33%	<u>XM_006715630.4</u> <u>AC073343.6</u> XM_024250011.1
	Orycteropus afer afer zinc fing	er protein 316 (ZNF316)	. <u>mRNA</u>		150	150	100%	2e-32	84.67%	XM_007942750.1
PREDICTED:     Predicted:     Predicted:	Ochotona princeps zinc finger pistrellus genome assembly.	r protein 316 (ZNF316), r chromosome: 5	<u>mRNA</u>		145 145 145	145 145	98% 100%	7e-31 7e-31	84.35% 84.11%	<u>XM_012930995.1</u> LR862361.1

Fig.9b Examining these reads revealed that they are only found in humans and apes. This is therefore also clear evidence that there are Human/Hominid-derived contamination in **SRX6893139**.

Description	Homo sapiens chromoson	ne 9, clone hRPK.2	02_H_3, complet	Percent Identity	E	value			Query Coverage
Molecule type	nucleic acid			to			to		to
Query Length	187174								
Other reports	Distance tree of results M	SA viewer 🔞							Filter Reset
Descriptions	Graphic Summary	Alignments							
Sequences	producing significant a	lignments		Dowr	nload ~	Ma	nage Co	olumns	✓ Show 100 ✔ 🔮
🗹 select all	100 sequences selected							Graphi	ics Distance tree of results
		Descriptio	n		Max 1 Score S	otal Quer core Cove	y E r value	Per. Ident	Accession
SRX689315	57				278	278 0%	8e-70	100.00%	SRA:SRR10168374.7906491.2

Fig.10a One read is also recovered from <u>SRX6893157</u>. From a query sequence only 187174nt in length.

		PREDIC	CTED: Homo sapiens formin binding protein 1 (FNBP1), transcript variant X13, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	005251824	.2	
		PREDIC	CTED: Homo sapiens formin binding protein 1 (FNBP1), transcript variant X4, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	011518402	.1	
		PREDIC	CTED: Homo sapiens formin binding.protein 1 (FNBP1), transcript variant X3, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	011518401	1	
		Homo s	sapiens formin binding.protein 1.(ENBP1). RefSeqGene on chromosome 9	278	278	100%	6e-71	100.00	% <u>NG</u>	033946.1		
		Homo s	sapiens cDNA FLJ13619 fis, clone PLACE1010926, weakly similar to HYPOTHETICAL 72.2 KD PROTEIN C12C2.05C IN CHROMOS(	278	278	100%	6e-71	100.00	% <u>AK(</u>	023681.1		
		Human	DNA sequence from clone RP11-138E2 on chromosome 9q34.11-34.3, complete sequence	278	278	100%	6e-71	100.00	% <u>AL1</u>	136141.13		
		Homo s	sapiens formin-binding.protein 17. (FBP17) mRNA_partial cds	278	278	100%	6e-71	100.00	% <u>AF2</u>	265550.1		
		Homo s	sapiens chromosome 9, clone hRPK.202_H_3, complete sequence	278	278	100%	6e-71	100.00	% <u>AC(</u>	006241.1		
		Homo s	sapiens KIAA0554 mRNA for KIAA0554 protein	278	278	100%	6e-71	100.00	% <u>AB</u>	011126.1		
		PREDIC	CTED: Nomascus leucogenys formin binding protein 1 (ENBP1), transcript variant X18, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818029	.1	
		PREDIC	CTED: Nomascus leucogenys formin binding protein 1 (FNBP1), transcript variant X17, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818028	1	
		PREDIC	CTED: Nomascus leucogenys formin binding protein 1 (FNBP1), transcript variant X16, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818027	1	
Dese	criptior	ı	gnl SRA SRR10168374.7906491.2 7906491 (Biological)									
Mole	ecule ty	/pe	dna									
Que	ry Leng	gth	150									
Oth	er repo	rts	Distance tree of results MSA viewer 😵									
~	Sciuru	is caro	linensis genome assembly, chromosome: 16		17	4 1	74 9	5% 9	e-40	88.81%	LR738606.1	
~	<u>Sciuru</u>	is caro	linensis genome assembly, chromosome: 14		17	4 1	74 9	5% 9	e-40	88.81%	LR738604.1	
~	<u>Sciuru</u>	is vulg	aris genome assembly, chromosome: 15		16	9 10	69 9	5% 4	e-38	88.11%	LR738626.1	
✓	PRED	ICTED	: Loxodonta africana formin binding protein 1 (FNBP1), transcript variant X2, mRNA		13	5 13	35 7	4% 4	e-28	88.50%	XM 010587	<u>565.2</u>
~	PRED	ICTED	): Loxodonta africana formin binding protein 1 (FNBP1), transcript variant X1, mRNA		13	5 13	35 7	4% 4	e-28	88.50%	XM_023544	<u>839.1</u>

Fig.10b This particular sequence is only found in humans—indicating that even the <u>SRX6893157</u> dataset was contaminated by material of human origin.

Description Molecule type Query Length Other reports	Percent Identity		Eva	alue	to		Query Coverage to Filter Reset		
Descriptions	Graphic Summary	Alignments							
Sequences p	oroducing significant a	alignments	Dov	wnloa	4 ×	Ма	nage (	Columns	✓ Show 100 ✔ ②
select all	100 sequences selected							Grap	hics Distance tree of results
		Description		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
SRX775676	<u>15</u>			279	774	0%	4e-69	100.00%	SRA:SRR11119763.129105044.1
SRX775676	5			279	375	0%	4e-69	100.00%	SRA:SRR11119763.106754018.1
SRX775676	<u>5</u>			279	279	0%	4e-69	100.00%	SRA:SRR11119763.75715814.1
SRX775676	5			279	510	0%	4e-69	100.00%	SRA:SRR11119763.47541416.1
SRX775676	<u>5</u>			279	279	0%	4e-69	100.00%	SRA:SRR11119763.40703161.1
SRX775676	5			279	768	0%	4e-69	100.00%	SRA:SRR11119763.38948039.1

Fig.11a The presence of Reads from Somatic Chlorocebus aethiops in <u>SRX7756765</u> confirms the identity of the Cercopithecinae reads there.

Desci	ription	gnl SRA SRR11119763.106	754018.1 106754018	(Biological)		to		to				to	
Mole	cule type	dna											
Quer	y Length	151									Filter	Rese	t
Othe	r reports	Distance tree of results	2										
				-									
De	scriptions	Graphic Summary	Alignments	Taxonomy									
Se	quences p	producing significant a	lignments			Download	<b>:</b> ~	Manage	Colum	ns ~	Show	1000 🗸	0
	select all	1000 sequences selected						<u>Genl</u>	<u>Bank</u>	Graphic	<u>s Dista</u>	nce tree of re	<u>sults</u>
			Descriptio	n			Max	Total	Query	E	Per.	Accessio	n
	Oblemente	us authings RAO slags OLIOSO	7004 from observed and	- C			Score	Score	100%	value	100.00%	AC241406.2	
	Chlorocebu	us aethiops BAC clone CH252-2	1451 10 from chromosom	<u>e o, complete sequence</u>	uence		279	690	100%	2e-71	100.00%	AC238953.2	
	Macaca mi	ilatta Y Chr BAC CH250-11 113 (	Children's Hospital Oak	and Research Institute Rh	esus maraque	e Adult Male BAC	279	1513	100%	2e-71	100.00%	AC240710.4	
	Chlorocebu	us aethiops BAC clone CH252-3	371E9 from chromosom	e 5. complete sequence	oodo madaqa		279	1124	100%	2e-71	100.00%	AC239563.3	
	Chlorocebu	us aethiops BAC clone CH252-4	461K13 from chromosor	ne 13, complete seguence			279	629	100%	2e-71	100.00%	AC239354.3	
	Chlorocebu	us aethiops BAC clone CH252-1	138D20 from chromosor	ne 13, complete sequence			279	493	100%	2e-71	100.00%	AC239463.3	
	Chlorocebu	us aethiops BAC clone CH252-4	17L1 from chromosom	e 6, complete sequence			279	472	100%	2e-71	100.00%	AC239275.3	
	Chlorocebu	us aethiops BAC clone CH252-6	2B18 from chromosom	<u>e 4, complete sequence</u>			279	510	100%	2e-71	100.00%	AC239442.2	
	Eukaryotic sy	nthetic construct chromosome 1	18				268	1.545e+05	100%	4e-68	100.00%	CP034496.1	
	Eukaryotic sy	nthetic construct chromosome 1	<u>16</u>				268	1.008e+05	100%	4e-68	98.68%	<u>CP034494.1</u>	
	Eukaryotic sy	nthetic construct chromosome 1	<u>19</u>				263	47303	100%	2e-66	98.01%	CP034522.1	
	Eukaryotic sy	nthetic construct chromosome 1	<u>13</u>				263	1.935e+05	100%	2e-66	99.31%	CP034516.1	
	Eukaryotic sy	nthetic construct chromosome \	Ľ				263	1.175e+05	100%	2e-66	99.31%	CP034510.1	
	Eukaryotic sy	nthetic construct chromosome 2	<u>21</u>				263	55262	100%	2e-66	98.01%	CP034500.1	
	Eukaryotic sy	nthetic construct chromosome 2	20				263	88148	100%	2e-66	99.31%	<u>CP034499.1</u>	
	Eukaryotic sy	nthetic construct chromosome 1	<u>19</u>				263	47481	100%	2e-66	98.01%	CP034497.1	
	Eukaryotic sy	nthetic construct chromosome 1	17				263	74732	100%	2e-66	98.01%	CP034495.1	
	Eukaryotic sy	nthetic construct chromosome 1	<u>15</u>				263	1.384e+05	100%	2e-66	98.01%	CP034493.1	
	Eukaryotic sy	nthetic construct chromosome 1	<u>14</u>				263	1.910e+05	100%	2e-66	99.31%	CP034492.1	
	Eukaryotic sy	nthetic construct chromosome 1	<u>13</u>				263	1.935e+05	100%	2e-66	99.31%	CP034491.1	
~	Eukaryotic sy	nthetic construct chromosome 2	22				257	22814	100%	9e-65	97.35%	CP034501.1	
~	<u>Human gamr</u>	maherpesvirus 4 isolate HKNPC	<u>260, partial genome</u>				248	248	100%	5e-62	96.03%	MH590571.1	
~	<u>Human gamr</u>	maherpesvirus 4 isolate HKHD4	40, partial genome				248	248	100%	5e-62	96.03%	<u>MH590409.1</u>	
~	Uncultured ba	acterium clone LM0ABA39ZH04	FM1 genomic sequence	1			246	246	96%	2e-61	97.24%	EU058958.1	
	Synthetic con	struct Pan troglodytes LINE-1L	1Pt retrotransposon tag	ged with EGFP sequence			241	241	100%	9e-60	95.36%	KF661301.1	
$\sim$	Uncultured or	rganism clone VC1C968TR gen	nomic sequence				241	241	96%	9e-60	96.55%	<u>GQ879596.1</u>	

Fig.11b the sequences from the BLAST hits indicate that they were unique to the family Cercopithecinae. Confirming Primate origin.

#### Analyzing the extent of contamination.

As the Specific BLAST analysis confirmed significant level of Human-derived contamination in all samples positive for SARS-CoV-2 related Coronaviruses, The TRACE result can therefore be

trusted for the analysis on the extent of contamination.

The 32nt Krona Trace system is used for elucidating the ratio of different taxa within a sample. As Specific BLAST analysis confirmed the significant presence of Human and Primate derived Genetic material--The most basal group of primates detected in all Coronavirus-positive samples belong to Catarrhini—or Humans, Apes and Old-World Monkeys. Therefore, Trace classification results that can be classified into sister nodes of Catarrhini should be considered as Contamination by Primate-derived material.

Since Catarrhini is under Simiiformes; Haplorrhini; Primates; Euarchonta; Euarchontoglires and Manis is under Pholidota; Laurasiatheria, If a read is TRACEd down to Catarrhini, it can not be from a Pangolin, and it will have to be from a Primate-derived source—Contamination by material from the lab.



Fig. 12 Family tree of mammals, Including the position and classification of Primates in the lineage of Mammalia.

Table 3a Ratios of Hominid-traced reads to Pangolin-traced reads in the SRA datasets that contained reads of the GD- Pangolin-CoV sequence, and had Hominid reads.

Accession	and	Primate	Total traced Kbps	Ratio of	Virus
date		classification and	to Manis Javanica	Primate to	classification
		total traced Kbps	(Pangolin)	Pangolin	and amount of
					reads by Kbps
SRX7756769		Homo sapiens	15401134	0.35	Bat SARS-like
18-Feb-2020		5457929			coronavirus
					2Kbp
					Wuhan seafood
					market
					pneumonia
					virus 2Kbp
SRX6893139		Homo sapiens	5301351	0.0926	Pangolin
20-Sep-2019		491120			coronavirus
					2Kbp
SRX6893157		Catarrhini	1889448	0.34	N/D***
20-Sep-2019		644546			
SRX6893156		Homo sapiens	4765461	0.01719	Pangolin
20-Sep-2019		81948			coronavirus
					2Kbp
SRX6893155		Homininae	525801	6.7214	Pangolin
20-Sep-2019		3534150			coronavirus
					5Kbp
SRX6893154		Hominoidea	2232008	0.159	Pangolin
20-Sep-2019		356003			coronavirus
					154Kbp
SRX6893153		Homo sapiens	3110158	0.05214	Pangolin
20-Sep-2019		162180			coronavirus
					41Kbp

\*\*\*: No trace result on Coronaviruses, despite claimed reads from [3]

Table 3b Ratios of Primate-traced reads to Coronavirus-traced reads in the SRA datasets that contained reads claimed to be traced to of the GD- Pangolin-CoV sequence, and lacked Hominid reads.

Accession and date	Primate classification	Virus	Ratio of virus
	and reads (in Kbp)	classification and	reads to
		reads	Primate reads
SRX7756766	Cercopithecidae 3116;	Betacoronavirus	0.000642
18-Feb-2020	BLAST to Macaca	2Kbp **	
	Mulatta		
SRX7756762	Catarrhini 2831;	Nidovirales OKbp	0.000530
18-Feb-2020	BLAST to Chlorocebus	Claimed	
	sabaeus	10x150bp reads	
SRX7756765	Cercopithecinae 11339	N/D***	N/A
22-Apr-2020	BLAST to Chlorocebus		
	Aethiops		
SRX7732094	N/A*	Pangolin	N/A*
15-Feb-2020		coronavirus	

\*: No non-coronavirus reads available in the dataset with a total of 2,633 reads, making analysis impossible.

\*\*: No claimed reads from [2]

\*\*\*: Claimed 8 reads from [10]

# DISCUSSIONS

#### The extent of contamination in the pangolin sequencing datasets

As the samples were supposed to be pangolin lung tissue, which will neither contact with nor be contaminated by non-pangolin derived mammalian tissues when still inside the animal, any non-pangolin mammalian reads within such a dataset can only be introduced to the sequencing process after the sample itself have been taken and brought into a lab.

As the classification Catarrhini itself is phylogenetically very deep down the Primate line which is itself distinguished from the Pangolin line at a very basal node (Boreoeutheria), and since we have already confirmed that the Primate line in PRJNA573298 traces mostly to humans by using Specific BLAST analysis, (SRX6893157, the only one of the claimed coronavirus read dataset that gives a classification just down to Catarrhini, contained 213 full length 100% matches to the Human Mitochondrial reference genome alone, which is only 16569 bp in length. All other datasets gives definitive TRACE mapping to Homo Sapiens and contained distinct 100% matched reads to even very small parts of the Human genome.), We can deduce the extent of contamination of the PRJNA573298 dataset by Primate-related materials as from a minimum of 1.6% to as high as 87% by sample mass—using the ratio of Primate reads to Pangolin reads on TRACE. Such high level of contamination with Primate-derived material is unacceptable for a sample that was supposed to be Lung tissue. And therefore, the virome data of such samples in PRJNA573298 no longer reflects the original virome of the animal, and an potential "novel" reads from these contaminated samples may have been from in-lab contamination instead.

#### Deducing the dynamic of contamination in PRJNA607174

Of all 7 PRJNA607174 datasets, only <u>SRX7756769</u> and <u>SRX7756762</u> is claimed by Xiao et. Al to contain SARS-CoV-2-like reads. However, TRACE results revealed low level of contamination by Cercopithecidae (Old World Monkey) reads across all the samples. In particular, the <u>SRX7756762</u> dataset contained definitive mappings to Chlorocebus sabaeus, or African Green Monkey, while <u>SRX7756766</u> which contained 2Kbp unclaimed reads of Betacoronaviruses on TRACE, contained 100% full-length definitive mappings to Macaca Mulatta that may also be mapped to Chlorocebus Aethiops and Homo Sapiens.

<u>SRX7756769</u> genetically resembles other samples in PRJNA573298, in both the kind of contamination and the extent of contamination. It contained an large excess of homo sapiens reads in levels similar to the contaminated samples in PRJNA573298.

From the method section of Lam et.al, we knew that they have performed Virus isolation using VERO E6 cells—Species Chlorocebus Sabaeus on one of the samples that have a positive PCR test for coronaviruses. The low level of contamination by Cercopithecidae-related reads in all the samples in PRJNA607174 except for <u>SRX7756769</u> itself support the possibility that <u>SRX7756769</u> is the first sample to be sequenced, and it happens before the lab begun using VERO E6 cells in the experiment. They then isolated the virus from the contaminated <u>SRX7756769</u> in VERO E6 cells, characterized it but did not sequence it, and this cell culture material then contaminated <u>SRX7756762</u> and possibly <u>SRX7756766</u>, resulting the 10 reads in <u>SRX7756762</u> and the 2Kb batacoronavirus reads in <u>SRX7756766</u>.

#### The exact nature of <u>SRX7732094</u> needs to be further scrutinized.

The P2S dataset, SRX7732094, displays very unusual property when compared to other Datasets under the same BioProject. It is the only dataset with all Non-coronavirus reads being filtered out, and contained too little spots for it to be an ILLUMINA NextSeq 550 run. Furthermore, it was the only dataset that did not contain metadata with either an isolation source or a Library prep procedure, other than "This dataset contains coronavirus-like sequence reads, based on BLAST search."

Such a strange designation and the fact of the dataset being heavily filtered, Raises problems on whether such a dataset is an actual BioSample at all. If this sample is really as claimed by Lam et. Al, Why the dataset have to be put through such heavy filtering when the other sequencing runs was clearly not filtered as severely as this dataset? Why there was no BioSample metadata on either Biomaterial provider, Source Tissue or Collector when all other Sequencing runs clearly provided such metadata information?

Unless the complete, unfiltered sequencing reads are made available on **SRX7732094**, and the rest of **PRJNA696875**, this Dataset can not be considered to be a real, reliable sample, and it must be excluded as "evidence" of a SARS-CoV-2-like virus infecting

pangolins in GuangDong, 2019.

Table 4	Sequencing	runs in	PRJNA696875,	Accession	number,	BioSample,	Content
and desi	gnation						

Accession	Size	Non-Coronavirus	Source	Virus	Design
number and		reads?	Tissue	Designation:	
date			Provider	GD or GX?	
			and		
			Collected		
			by		
SRX7732094	2,633	No	N/A	GD	This dataset
15-Feb-2020					contains
					coronavirus-like
					sequence
					reads, based on
					BLAST search.
SRX7732093	470,344	Yes	Intestine	GX	NEBNext Ultra
15-Feb-2020			Yanling Hu		II DNA Library
			Wuchun		Prep Kit, paired
			Cao		sequencing
					data has been
					integrated.
SRX7732092	340,661	Yes	Lung	GX	NEBNext Ultra
15-Feb-2020			Yanling Hu		II DNA Library
			Wuchun		Prep Kit, paired
			Cao		sequencing
					data has been
					integrated.
SRX7732091	416,659	Yes	Intestine	GX	NEBNext Ultra
15-Feb-2020			Yanling Hu		II DNA Library
			Wuchun		Prep Kit, paired
			Cao		sequencing
					data has been
					integrated.
SRX7732090	520,254	Yes	Lung	GX	NEBNext Ultra
15-Feb-2020			Yanling Hu		II DNA Library
			Wuchun		Prep Kit, paired
			Cao		sequencing
					data has been

					integrated.
SRX7732089	19,607,536	Yes	Blood	GX	Ion Total
15-Feb-2020			Yanling Hu		RNA-Seq Kit v2
			Wuchun		
			Cao		
SRX7732088	4,550,437	Yes	lung and	GX	Ion Total
15-Feb-2020			intestine		RNA-Seq Kit v2
			Yanling Hu		
			Wuchun		
			Cao		

By closely examining the P2V dataset, SRX7732088, which claimed to be a culture sample in VERO E6 cells, Chlorocebus Sabaeus, the exact viral load in-culture when compared to Cellular mRNA can be deduced by dividing the total identifiable coronavirus signal to the total identifiable Primate signal within the dataset, 6943Kbp/451932Kbp, which correspond to 0.01536:1 Viral RNA to Cellular RNA.

This places the viral loads on the other datasets with Coronavirus-like reads from GD well within the threshold expected from cell culture contamination of the sequencing samples—including the samples in PRJNA607174.

### Potential breach of data availability statement by Xiao et al.[2]

Sequence data that support the findings of this study have been deposi	ited in GISAID with the accession numbers EPI_	ISL_410721.	Raw data of RNAseq are available
from the NCBI SRA under the study accession number PRJNA607174.			

Fig 13. The Data Availability Statement of Xiao et al.

In the Data availability statement, the "Raw data of RNAseq" are clearly stated to be deposited under PRJNA607174. However, only 2 of the "Extended Data Table S3" datasets actually matches the datasets deposited on PRJNA607174. The other 7 datasets were completely unavailable. And the actual deposited datasets on PRJNA607174 does not match what have been claimed by Extended Data Table S3. As the RNA-seq Raw data was stated to be available within PRJNA607174, the failure to publish all the claimed data constitute a breach of the Data Availability statement on the article. Unless such datasets are published and independently examined, All such claimed reads from the strangely unpublished datasets can not be trusted as evidence of a SARS-CoV-2-like virus infecting pangolins in GuangDong, 2019.

#### Identifying the Etiological agent of the GuangDong 2019 incident.

By using an approach of both SRA TRACE analysis and specific BLAST Analysis, We have uncovered the fact that all samples that does not Contain confirmed Human-derived material, also lacked Claimed reads of a SARS-CoV-2 like virus that can be confirmed using NCBI Trace. All samples with claimed or traced reads of Coronaviruses in general, contained confirmed primate reads with the lowest common phylogenetic node Catarrhini. Samples that does not give a TRACE result on primate-derived material all lacked identifiable or claimed coronavirus reads.

This strongly imply that the Coronavirus-like reads are associated with human/Primate-sourced contamination material.

Most importantly, of all dead pangolins being sampled in the studies, only 9 out of a total of 29 Analyzable samples/datasets contained TRACEd or Claimed Coronavirus reads—despite all dead pangolins displayed similar symptoms in captivity. This imply that the alleged pangolin coronavirus is not the Etiological agent of the death of the pangolins being sampled in the studies. This is further supported by the fact that 4 out of 10 lung samples in PRJNA573298 and 4 out of 7 lung samples in PRJNA607174 lacked any claimed or TRACEd coronavirus reads—despite the same symptoms displayed and similar date of death.

In order to establish the Etiological agent of the dead pangolins in the single GuangDone Accident that leads to the sampling and studies. A full virome TRACE analysis is conducted on the available samples for the determining of the exact etiological agent.

#### Extended Data Table S1

#### Full virome TRACE results of all Analyzable datasets of the GD pangolin incident

	Mammarenavirus	Nairoviridae	Murine respirovirus	Flaviviridae	Nidovirales	Rubulavirus	Nonanavirus	Peribunyay	Amigovirus	Siphoviridae	Siphoviridae	Pahexavir
SRX6893158	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	No	No
SRX6893157	Yes	Yes	No	No	Claimed	No	No	Yes	No	No	No	No
SRX6893156	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes
SRX6893155	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893154	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893153	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	No
SRX6893152	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No
SRX6893151	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	No
SRX6893150	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
SRX6893149	Yes	Yes	No	No	No	No	No	No	No	No	Yes	No
SRX6893148	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No
SRX6893147	Yes	Yes	"Respirovirus"	Yes	No	No	Yes	No	Yes	No	No	No
SRX6893146	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
SRX6893145	Yes	Yes	No	No	No	No	No	No	No	No	No	No
SRX6893144	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
SRX6893143	Yes	Yes	No	No	No	No	No	No	No	No	No	No
SRX6893142	Yes	Yes	No	No	No	No	No	Yes	Yes	No	No	No
SRX6893141	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No
SRX6893140	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No	No
SRX6893139	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893138	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No
SRX7756766	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No
SRX7756765	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756764	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756763	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756762	No	No	Yes	No	Claimed	Yes	No	No	No	No	No	No
SRX7756761	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756769	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No

A full Virome TRACE result suggest all the dead pangolins were infected by either Mammarenaviruses or Murine Respirovirus, or both. Including both samples that contained Claimed Or TRACEd Coronavirus reads and the samples that didn't.

Murine Respirovirus and Mammarenaviruses co-infect 7 out of 29 Available Analyzable datasets, while None of the 29 datasets lacked both—indicating that both viruses were prevalent in the location where the pangolins were captive at The Guangdong Wildlife Rescue Center.

Symptoms of Murine Respirovirus in animals resembles that of SARS-CoV-2 in humans—It forms massive Syncytiums in Eukaryotic cells, suppresses the immune system and causes secondary bacterial infections. The virus causes necrosis of Lung tissue in 5 days, with similar inflammation and immunopathological effects in the lung tissues of infected animals [5]—creating the histopathological effect as reported by Xiao et al.

It should be worth pointing out that the only examined lung tissues were examined by Xiao et al. And all Lung tissue samples examined by Xiao et.al contained Reads from the Murine Respirovirus.

Similarly, Mammarenaviruses are also known to cause multi organ, lethal[7] infections, characterized by endothelial pathology and swelling of internal organs. [6] All of which were Symptoms reported in the incident. As these samples were not examined Histopathologically by

either the authors of [4] nor by any of the authors of any other article who have used the datasets/samples, leaving the only mean of elucidating the cause of death being the observed symptoms and the coarse examination of the organs during sampling. Mammarenavirus infection therefore remains the most likely cause of death of the Murine Respirovirus Negative samples in the available datasets.

#### Is the "GD pangolin CoV" really a virus of the pangolin?

The only examination of the binding affinity of the GD pangolin CoV RBD to different animal receptors was done by Xiao et al [2], which performed molecular dynamic simulation of the RBD docking to the Human ACE2 receptor, The Civet ACE2 receptor and the pangolin ACE2 receptor. If the RBD of GD pangolin CoV in deed evolved in pangolins, we should expect the binding affinity of the RBD toward the pangolin ACE2 receptor to be the highest binding affinity returned from the examination.

However, neither the GD pangolin CoV RBD, nor the RBD of SARS-CoV-2 which is highly similar, produced a higher binding affinity to the pangolin ACE2 receptor than to the human ACE2 receptor, and both binds the Human ACE2 receptor with the highest affinity across all 3 animal species (Human, Civet, Pangolin) examined.

This fact argues strongly against the RBD residues of the GD pangolin CoV being evolved in pangolins, and instead favoring the RBD and the virus being the result of a passage experiment of a possible virus of pangolin origin (The GX/P2V virus was isolated and passaged in VERO E6 cells during it's collection in 2017) in Primate-derived cell lines.

There are only 2 locations of Biological sample storage in GuangDong, the Guangdong Institute of Applied Biological Resources and the China National GeneBank.

As all Credible (Non-filtered and contained analyzable Non-Coronavirus reads) samples were collected in a single incident from the GuangDong Wildlife Rescue Center[1][4][2], which the initial sample collection and storage was carried out by the Guangdong Institute of Applied Biological Resources[4], this experimental culture likely contaminated the GD pangolin samples during their initial collection or Storage, Either by the lab worker doing the initial sampling, or during their storage in the facility.

Epidemiology analysis of SARS-CoV-2 and related viruses argues strongly against the existence of a Coronavirus with the claimed RBD residues and sequence similarity in or near the GuangDong Wildlife Rescue Center at the time and date of the incident and the collection of the samples.

The earliest collection date of the GD pangolin CoV available, MP789, GenBank MT084071.1, is displayed at 29 March 2019.

Since the original location of the animals and samples in question was inside the GuangDong Wildlife Rescue Center which is neither a certified Biosafety Laboratory nor possessed adequate

PPE when handling the animals, from the Simulation results by Xiao et al[2] and the observed high human transmissibility of SARS-CoV-2 which had a very similar RBD, Should the GD pangolin CoV genuinely exists at that date and within the unprotected GuangDong Wildlife Rescue Center, It would almost certainly infect one to multiple On-site workers (Rescue workers which lacked either the Biosafety training or the adequate PPEs required to handle tissues or animals infected with a virus as characterized by the GD pangolin CoV papers) in the GuangDong Wildlife Rescue Center, and caused a SARS-level epidemic in GuangDong 2013 beginning in or around April 2019. However, no such epidemic was recorded, nor there have been any virus that genetically resembled the GD pangolin CoV sequence (which is only 90% similar to SARS-CoV-2) being isolated in humans anywhere in the world even till today.

Nor there is a possibility that the current SARS-CoV-2 pandemic may have stemmed from the 29 March incident with the GD pangolin CoV, since the estimated time of divergence between the current SARS-CoV-2 genome to the GD pangolin CoV Genome was estimated to be at least 100 years ago , ranging from 1851 [1730,1958] to 1877 [1746,1986] [8], for a genome that is only 90% similar to SARS-CoV-2 and possessed significant difference in the sequence and composition of the viral proteins they encodes.

As the Earliest time of discovery and the incident on the GD pangolin CoV is no earlier than the beginning of Year 2019, The time between the incident and the first isolate of SARS-CoV-2 is far too short for GD pangolin CoV incident to be involved in the formation of the current SARS-CoV-2 pandemic, since even the neutral sites on the RBD itself would have taken more than 19.8 years to drift/evolve into what we seen today on the actual SARS-CoV-2 genome. [9]

# Conclusions

The Extreme lack of transparency and the sheer level of contamination from the original samples, the lack of epidemiological evidence of it's existence at the location of it's collection, and the receptor binding affinity of the Viral RBD itself indicating it as not being evolved nor adapted in pangolins, all strongly argue against the existence of a SARS-CoV-2 like virus infecting pangolins captive in GuangDong at 2019.

Moreover, it suggests that the GD pangolin CoV exists only as a culture in Primate-derived cells within the lab/facility used for the initial collection and/or storage of the samples of the pangolins in question, raising important issues on the serial passage Gain-Of-Function research of viral pathogens.



Figure 14. A cartoon diagram of contamination in sequencing experiment leading to false results and false "discoveries".

# Note as in 2020/7/23

A recent Dataset, seemingly unrelated to the Xiao et.al Nature dataset, <u>SRX8582289</u>, appeared under <u>PRJNA607174</u>. This dataset seems to be newly sequenced, and it was not referred in [2].

Table S2: TRACE analy	sis result of the	SRX8582289	dataset.

Accession number and	Primary	Mamr	nalian	Primate-related	results	Identification	of
registration date	Trace	results	and	in Krona and rea	ad size	"Coronaviridae	e"
	percenta	ge		by Кbp	as by Trace and		
						total read size	
SRX8582289	Manis ja	vanica: <b>43.</b>	.52%	Catarrhini 98913		Pangolin	
22-Jun-2020						coronavirus 79	92

Nevertheless, in-depth analysis revealed significant amount of contamination from the Human genome, with ratio of Virus to cell=0.8%.

Description	Homo sapiens BAC clone R	P11-460N20 from 7, complete sequ	Percent Identity	E va	lue			Query Coverage				
Molecule type	nucleic acid	to			to		to					
Query Length	203396											
Other reports	Distance tree of results M						Filter Reset					
Descriptions	Graphic Summary	Alignments										
Sequences producing significant alignments Download × Manage Columns × Show 100 • 0												
🗹 select all	100 sequences selected						<u>Grap</u>	hics Distance tree of results				
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession				
SRX85822	<u>89</u>		278	278	0%	8e-69	100.00%	SRA:SRR12053850.88444297.1				
SRX85822	<u>89</u>		278	402	0%	8e-69	100.00%	SRA:SRR12053850.83916175.2				
SRX85822	<u>89</u>		278	388	0%	8e-69	100.00%	SRA:SRR12053850.83916175.1				
SRX85822	<u>89</u>		278	278	0%	8e-69	100.00%	SRA:SRR12053850.82221130.1				
SRX85822	<u>89</u>		278	278	0%	8e-69	100.00%	SRA:SRR12053850.71234261.2				
SRX85822	<u>89</u>		278	278	0%	8e-69	100.00%	SRA:SRR12053850.71234261.1				
SRX85822	<u>89</u>		278	5169	2%	8e-69	100.00%	SRA:SRR12053850.51889132.2				
SRX85822	<u>89</u>		278	7268	3%	8e-69	100.00%	SRA:SRR12053850.26027930.2				
SRX85822	<u>89</u>		278	5671	2%	8e-69	100.00%	SRA:SRR12053850.21554419.1				
SRX85822	<u>89</u>		278	278	0%	8e-69	100.00%	SRA:SRR12053850.13271287.2				
SRX85822	<u>89</u>		278	4760	1%	8e-69	100.00%	SRA:SRR12053850.62042.2				
SRX85822	<u>89</u>		276	276	0%	3e-68	100.00%	SRA:SRR12053850.82221130.2				

Descriptio	n gni SRA SRR12053850.82221130.2 82221130 (Biological)	to	1	to				to	
Molecule t	ype dna								
Query Len	gth 150						Filte	Res	et
)ther repo	Distance tree of results 😧								
Descrip	tions Graphic Summary Alignments Taxonomy								
Seque	scription gnl SRA SR12053850.82221130_282221130_(Biological) lecule type dna ery Length 150 Ter reports Distance tree of results Bescriptions Graphic Summary Alignments Taxonomy Descriptions Graphic Summary Alignments Taxonomy Sequences selected <u>Consecution and Consecution and</u>								
🗹 sele	ct all 59 sequences selected			<u>GenB</u>	<u>ank</u>	Graphic	<u>s Dista</u>	nce tree of re	sult
	Description		Max Score	Total Score	Query Cover	E value	Per. Ident	Accessio	on
Mo	mo sapiens general transcription factor IIi pseudogene 14.(GTF2IP14) on chromosome 7		276	276	99%	3e-70	100.00%	NG_043494	.1
<u>Ho</u>	mo sapiens FOSMID clone ABC13-48840700E15 from chromosome 7, complete sequence		276	276	99%	3e-70	100.00%	AC242196.4	
<u>Но</u>	mo sapiens BAC clone RP11-460N20 from 7, complete sequence		276	276	99%	3e-70	100.00%	AC073210.8	
<u>Ho</u>	mo sapiens general transcription factor lli pseudogene 5 (GTF2IP5) on chromosome 7		270	270	99%	1e-68	99.33%	NG 026590	.1
Par Par	n troglodytes BAC clone CH251-340124 from chromosome 7, complete sequence		270	270	99%	1e-68	99.33%	AC185242.2	
Par	n troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence		270	270	99%	1e-68	99.33%	AC184799.2	
Par	n troglodytes BAC clone CH251-114G16 from chromosome 7, complete sequence		270	270	99%	1e-68	99.33%	AC183835.2	
	n troglodytes BAC clone CH251-2015 from chromosome 7, complete sequence		270	270	99%	1e-68	99.33%	AC174000.3	
<u>Ho</u>	mo sapiens BAC clone RP11-47909 from 7, complete sequence		270	270	99%	1e-68	99.33%	AC073107.7	
Par	n troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence		270	270	99%	1e-68	99.33%	AC148313.3	
hor	mo sapiens BAC clone CH17-99D2 from chromosome 4. complete sequence		226	226	99%	3e-55	93.96%	AC278002.1	
<u>Ho</u>	mo sapiens FOSMID clone ABC27-154113 from chromosome 4, complete sequence		226	226	99%	3e-55	93.96%	AC240529.1	
Par	n troglodytes BAC clone CH251-4D23 from chromosome 7, complete sequence		226	226	99%	3e-55	93.96%	AC148834.3	

<u>34.3</u> Eukaryotic synthetic construct chromosome 16 193 193 99% 3e-45 90.07% <u>CP034494.1</u> Eukaryotic synthetic construct chromosome 19 182 182 99% 5e-42 88.74% CP034522.1 Eukaryotic synthetic construct chromosome 19 182 182 99% 5e-42 88.74% <u>CP034497.1</u>

Fig. S1B: BLAST results returned only Homo Sapiens as 100% match. This indicate that the listed Catarrhini reads come from Homo Sapiens.

The significance of this particular dataset is yet unknown.

# Note as in 2020/12/26

Two Recent SRAs, SRX9714436 and SRX9714921, were recently deposited by the Guangdong Institute of Applied Biological Resources with a listed DOI connection to 10.1371/journal.ppat.1008421 [1]. Both samples have a depositor of LinMao Li, 2020-12-21 the same time as the specified BioProject registration date. Only one of the SRAs contained significant amount of Coronavirus-related reads.

Accession number and	Primary Mamma	lian	Primate-related results	Identification of			
registration date	Trace results	and	in Krona and read size	"Coronaviridae"			
	percentage		by Кbp	as by Trace and			
				total read size			
<u>SRX9714436</u>	Manis javanica: 3.14%	6	Homo sapiens 12332	Pangolin			
	Homo sapiens: 0.04%	,		coronavirus 3			
SRX9714921	Homo sapiens: 0.15%	, )	Homo sapiens 9923	N/D			

Table S3: TRACE analysis result of SRX9714436 and SRX9714921.

As expected by TRACE results, Reads that are 100% full-length uniquely matched to Homo Sapiens were obtained from SRX9714436 and SRX9714921.

>gnl|SRA|SRR13285085.43 43

CTATACAACAAACCCCCATGACACGACTTTACCTATGTAACAAACCTTCA

	select all 100 sequences selected					<u>GenB</u>	ank	Graph	<u>nics</u> D	istance t	ree of result
	Description		Common Name	Max Score	Total Score	Query Cover	E value	P	er. A ent	cc. Len	Accession
~	Homo sapiens BAC clone RP11-488C22 from 2, complete sequence	<u>e</u>	human	93.5	93.5	100%	5e-16	5 100	.00% 1	65351	AC019109.9
	Pan troglodytes BAC clone CH251-617M1 from chromosome unknow	<u>wn, complete sequence</u>	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.0	00% 1	88195	AC183921.2
	Pan troolodytes BAC clone CH251-564M21 from chromosome unkn	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.0	00% 2	201789	AC160021.3	
	Homo sapiens chromosome 5, BAC clone 89K19 (LBNL H179), co	<u>human</u>	87.9	87.9	100%	2e-14	98.0	00% 1	110312	AC005215.2	
	Homo sapiens chromosome 5 clone CTD-2332G20, complete seq	uence	<u>human</u>	87.9	87.9	94%	2e-14	100	.00% 1	46437	AC010489.4
	Eukaryotic synthetic construct chromosome Y		eukaryotic syn	84.2	168	90%	3e-13	3 100	00% 64	558752	CP034510.1
~	Pongo abelii BAC clone CH276-222M18 from chromosome 8. com	<u>plete sequence</u>	Sumatran ora	84.2	84.2	90%	3e-13	3 100	.00% 1	96788	AC206339.2
	Description	Common Na	me	S	Max T core S	Fotal C core C	over v	E alue	Per. Ident	Acc. Le	en Accessi
	Eukaryotic synthetic construct chromosome Y	eukaryotic synthetic construct		;	34.2	168 9	90% :	3e-13	100.009	645587	52 <u>CP03451</u>
~	Eukaryotic synthetic construct chromosome 15	eukaryotic synthetic construct			80.5	80.5 9	92%	4e-12	97.83%	825213	92 <u>CP03449</u>
~	Eukaryotic synthetic construct chromosome 14	eukaryotic synthetic construct		:	76.8	151 9	94%	5e-11	95.74%	882895	40 <u>CP03449</u>

	Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct	75.0	210	100%	2e-10	95.65%	84899006	CP034496.1
	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034516.1
	Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	98200793	CP034494.1
	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034491.1
~	Eukaryotic synthetic construct chromosome 20	eukaryotic synthetic construct	69.4	138	86%	8e-09	95.35%	68480253	CP034499.1
	Eukarvotic synthetic construct chromosome 19	eukarvotic synthetic construct	62.1	62.1	84%	1e-06	92.86%	64242768	CP034497.1

Fig.S2A: Homo Sapiens 100% full-length Unique matched read obtained from SRX9714436

->gni|SRA|SRR13285570.36 36 ATAGGGAAGTGTGGTACCAAGGAGCAATATTCAATACAGCAACCAGGAAG

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
C	Human DNA sequence from clone XX-DSH1_29E11, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	112322	CU041292.6
	Human DNA sequence from clone RP11-535B18 on chromosome 9, complete sequence	<u>human</u>	93.5	93.5	100%	5e-16	100.00%	126815	AL354931.13

Fig.S2B: Homo Sapiens 100% full-length Unique matched read obtained from <u>SRX9714921</u> A provided .fastq file was also found in <u>SRX9714436</u>. Analysis suggest significant presence of Homo Sapiens reads within this fastq file, similar to that of the Run itself.

NH:i:1	AS:i:	50	NM:i:0					
50M *	0	0	TGAAGGTTTGTTACATAGGT	AAAC	TCGTGTCATGGG	GGTTTGTTG	FATAG	*
NDX550397_RU	0:309:	H3F	WBGXH:1:11101:21446:1055	16	AC019109.9	87488	255	

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<	Homo sapiens BAC clone RP11-488C22 from 2, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	165351	AC019109.9
	Pan troglodytes BAC clone CH251-617M1 from chromosome unknown, complete sequence	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	188195	AC183921.2
	Pan troglodytes BAC clone CH251-564M21 from chromosome unknown, complete sequence	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	201789	AC160021.3
	Homo sapiens chromosome 5, BAC clone 89K19 (LBNL H179), complete sequence	<u>human</u>	87.9	87.9	100%	2e-14	98.00%	110312	AC005215.2
	Homo sapiens chromosome 5 clone CTD-2332G20, complete sequence	<u>human</u>	87.9	87.9	94%	2e-14	100.00%	146437	AC010489.4
	Eukaryotic synthetic construct chromosome Y	eukaryotic syn	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
	Pongo abelii BAC clone CH276-222M18 from chromosome 8, complete sequence	Sumatran ora	84.2	84.2	90%	3e-13	100.00%	196788	AC206339.2
	Human DNA sequence from clone RP11-987D21 on chromosome X, complete sequence	<u>human</u>	84.2	84.2	96%	3e-13	97.92%	55442	<u>BX119919.5</u>
	PREDICTED: Callithrix jacchus uncharacterized LOC118154814 (LOC118154814), ncRNA	white-tufted-e	82.4	82.4	94%	1e-12	97.87%	2719	<u>XR_004745075.1</u>
	Homo sapiens solute carrier family 26 member 3 (SLC26A3), RefSeqGene (LRG_683) on chromos.	. <u>human</u>	82.4	82.4	100%	1e-12	96.00%	44767	<u>NG_008046.1</u>
	Eukaryotic synthetic construct chromosome Y eukaryotic synthetic construct		8	34.2	168 9	10% 31	e-13 100.	00% 64558	3752 <u>CP034510.1</u>
<	Eukaryotic synthetic construct chromosome 15 eukaryotic synthetic construct		8	80.5	30.5 9	2% 4	e-12 97.8	3% 8252	1392 <u>CP034493.1</u>
	Eukaryotic synthetic construct chromosome 14 eukaryotic synthetic construct		7	76.8	151 9	4% 5	e-11 95.7	4% 88289	9540 <u>CP034492.1</u>
	Eukaryotic synthetic construct chromosome 18 eukaryotic synthetic construct		7	75.0	210 10	00% 2	e-10 95.6	5% 84899	0006 <u>CP034496.1</u>
	Eukaryotic synthetic construct chromosome 13 eukaryotic synthetic construct		7	71.3	71.3 9	4% 2	e-09 93.6	2% 96089	9878 <u>CP034516.1</u>
	Eukaryotic synthetic construct chromosome 16 eukaryotic synthetic construct		7	71.3	71.3 9	4% 2	e-09 93.6	2% 98200	0793 <u>CP034494.1</u>
	Eukaryotic synthetic construct chromosome 13 eukaryotic synthetic construct		7	71.3	71.3 9	4% 2	e-09 93.6	2% 96089	9878 <u>CP034491.1</u>
<	Eukaryotic synthetic construct chromosome 20 eukaryotic synthetic construct		6	59.4	138 8	6% 8	e-09 95.3	5% 68480	0253 <u>CP034499.1</u>
	Eukaryotic synthetic construct chromosome 19 eukaryotic synthetic construct		6	62.1 (	52.1 8	4% 1	e-06 92.8	6% 64242	2768 <u>CP034497.1</u>

Fig.S3: BLAST result of the read NDX550397\_RUO:309:H3FKWBGXH:1:11101:21446:1055 TGAAGGTTTGTTACATAGGTAAACTCGTGTCATGGGGGGTTTGTTGTATAG within the provided fastq file.

The read is 100% full-length uniquely matched to Homo Sapiens.

# REFERENCES

[1] Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?

Ping Liu , Jing-Zhe Jiang , Xiu-Feng Wan, Yan Hua, Linmiao Li, Jiabin Zhou, Xiaohu Wang, Fanghui Hou, Jing Chen, Jiejian Zou, Jinping Chen

Published: May 14, 2020

https://doi.org/10.1371/journal.ppat.1008421

[2] Xiao, K., Zhai, J., Feng, Y. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020). <u>https://doi.org/10.1038/s41586-020-2313-x</u>

[3] Lam, T.T., Shum, M.H., Zhu, H. *et al.* Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* (2020). <u>https://doi.org/10.1038/s41586-020-2169-0</u>

[4] Liu, P.; Chen, W.; Chen, J.-P. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). *Viruses* **2019**, *11*, 979.

[5] Inducible epithelial resistance improves survival of Sendai virus pneumonia in mice by both inactivating virus and preventing CD8+ T cell-mediated immunopathology

S. Wali, J. R. Flores, A.M. Jaramillo, D. L. Goldblatt, J. Pantaleón García, M. J. Tuvim, B. F. Dickey, S. E. Evans

doi: https://doi.org/10.1101/2020.01.30.917195

[6] Jorlan Fernandes, Renata Carvalho de Oliveira, Alexandro Guterres, Débora Ferreira Barreto-Vieira, Ana Claudia Pereira Terças, Bernardo Rodrigues Teixeira, Marcos Alexandre Nunes da Silva, Gabriela Cardoso Caldas, Janice Mery Chicarino de Oliveira Coelho, Ortrud Monika Barth, Paulo Sergio D'Andrea, Cibele Rodrigues Bonvicino, Elba Regina Sampaio de Lemos,

Detection of Latino virus (Arenaviridae: Mammarenavirus) naturally infecting Calomys callidus,

Acta Tropica, Volume 179, 2018, Pages 17-24, ISSN 0001-706X, https://doi.org/10.1016/j.actatropica.2017.12.003. (http://www.sciencedirect.com/science/article/pii/S0001706X17311749)

[7] Hemorrhagic Fever-Causing Arenaviruses: Lethal Pathogens and Potent Immune Suppressors Morgan E. Brisse1,2 and Hinh Ly2,\*

[8] Evolutionary origins of the SARS - CoV - 2sarbecovirus lineage responsible for the COVID-19 pandemicMaciej F Boni1\* , Philippe Lemey2\* , Xiaowei Jiang3, Tommy Tsan-Yuk Lam4, Blair Perry5, Todd Castoe5, Andrew Rambaut6 and David L Robertson7

[9] Xiaolu Tang, Changcheng Wu, Xiang Li, Yuhe Song, Xinmin Yao, Xinkai Wu, Yuange Duan, Hong Zhang, Yirong Wang, Zhaohui Qian, Jie Cui, Jian Lu, On the origin and continuing evolution of SARS-CoV-2, *National Science Review*, , nwaa036, <u>https://doi.org/10.1093/nsr/nwaa036</u>

[10] SARS-CoV-2-like viruses from captive Guangdong pangolins generate circular RNAs

Alexandre Hassanin 1 Huw Jones 2 Anne Ropiquet 2