repiSalud
Repositorio Institucional de Salud

This is the peer reviewed version of the following article:

CHARACTERISATION OF LGP2 COMPLEX MULTITRANSCRIPT SYSTEM IN HUMANS: ROLE IN THE INNATE IMMUNE RESPONSE AND EVOLUTION FROM NON-HUMAN PRIMATES

Jorge Martinez-Laso, Isabel Cervera, Marina S Martinez-Carrasco, Veronica Briz, Celia Crespo-Bermejo, Clara Sánchez-Menéndez, Guiomar Casado-Fernández, Montserrat Torres, Mayte Coiras, Characterisation of LGP2 complex multitranscript system in humans: role in the innate immune response and evolution from non-human primates, Human Molecular Genetics, Volume 34, Issue 1, 1 January 2025, Pages 11–20.

which has been published in final form at:
https://doi.org/10.1093/hmg/ddae155

1
Characterisation of LGP2 complex multitranscript system in humans: role in the
2
innate immune response and evolution from non-human primates

3

4 Jorge Martinez-Laso[1,*], Isabel Cervera[1], Marina S. Martinez-Carrasco[1,2], Veronica Briz[3],
5 Celia Crespo-Bermejo[3], Clara Sánchez-Menéndez[4,5,6], Guiomar Casado-Fernández[4,7],
6 Montserrat Torres[4,8], Mayte Coiras[4,8].
7 [1]Immunogenetics Unit. National Center of Microbiology, Instituto de Salud Carlos III,
8 28220 Madrid, Spain.
9 [2]Pediatrics Department. Hospital Universitario 12 de Octubre. Avda de Córdoba s/n
10 28041 Madrid
11 [3]Viral Hepatitis Reference and Research Laboratory, National Center of Microbiology,
12 Institute of Health Carlos III, Madrid, Majadahonda, Spain.
13 [4]Immunopathology and Viral Reservoir Unit, National Center of Microbiology, Instituto
14 de Salud Carlos III, Majadahonda, Madrid, Spain.
15 [5]PhD Program in Biomedical Sciences and Public Health, Universidad Nacional de
16 Educación a Distancia (UNED), Madrid, Spain.
17 [6]Hematology and Hemotherapy Service, Instituto Ramón y Cajal de Investigación
18 Sanitaria (IRYCIS), Hospital Universitario Ramón y Cajal, Madrid, Spain.
19 [7]PhD Program in Health Sciences, Faculty of Sciences, Universidad de Alcalá, Alcalá de
20 Henares, Spain.
21 [8]Biomedical Research Center Network in Infectious Diseases (CIBERINFEC), Instituto
22 de Salud Carlos III, Majadahonda, Madrid, Spain.
23 *Correspondence: Jorge Martinez-Laso, Immunogenetics Unit, National Center of
24 Microbiology Instituto de Salud Carlos III. Ctra Majadahonda-Pozuelo K2,2, 28220
25 Majadahonda, Madrid. e-mail: jmlaso@isciii.es; Tel.: 34918223751

26 **ABSTRACT**

27 Retinoic acid inducible gene I (RIG-I)-like receptors (RLRs), including RIG-I, MDA5
28 and LGP2, recognize viral RNA to mount an antiviral interferon (IFN) response RLRs
29 share three different protein domains: C-terminal domain, DExD/H box RNA helicase
30 domain, and an N-terminal domain with two tandem repeats (CARDs). LGP2 lacks
31 tandem CARD and is not able to induce an IFN response. However, LGP2 positively
32 enhances MDA5 and negatively regulates RIG-I signaling. In this study, we determined
33 the LGP2 alternative transcripts in humans to further comprehend the mechanism of its
34 regulation, their evolutionary origin, and the isoforms functionallity. The results showed
35 new eight alternative transcripts in the samples tested. The presence of these transcripts
36 demonstrated that the main mechanisms for the regulation of LGP2 expression are both
37 by insertion of introns and by the loss of exons. The phylogenetic analysis of the
38 comparison between sequences from exon 1 to exon 3 of humans and those previously
39 described in non-human primates showed three well-differentiated groups (lineages)
40 originating from gorillas, suggesting that the transspecies evolution has been maintained
41 for 10 million years. The corresponding protein models (isoforms) were also established,
42 obtaining four isoforms: one complete and three others lacking the C-terminal domain or
43 this domain and the partial or total He2 Helicase domain, which would compromise the
44 functionality of LGP2. In conclusion, this is the first study that elucidate the large
45 genomic organization and complex transcriptional regulation of human LGP2, its pattern
46 of sequence generation, and a mode of evolutionary inheritance across species.

47

48

49

50

51 **INTRODUCTION**

52 The human innate immune system through their pattern-recognition receptors 53 (PRR) present in macrophages and dendritic cells recognize pathogens associated 54 molecular patterns (PAMPS) in microbes that do not exist in humans (1-3).

55 Once the viral nucleic acids have been recognized, the PRRs launch a series of 56 pathways that lead to the production of type I interferons (IFNα, IFNβ, IFNω, IFNε or 57 IFNκ) and inflammatory cytokines (IL-6, TNFα and IL-12) that play an essential role in 58 virus clearance. Of the two receptor systems that initiate the interferon-dependent 59 immune response, the TLRs, (Toll-like receptors) and the RLRs (RIG-I-like receptors) 60 (4-11), the latter are still largely unknown in terms of their genetic, immunological and 61 evolutionary characteristics.

62 The RLR family of cytoplasmic RNA helicases DExH/D detects foreign double-63 stranded (ds) RNA and is composed of three members RIG-I (retinoic acid-inducible gene 64 I or DDX58), MDA5 (retinoic acid-inducible melanoma differentiation-associated gene 65 5 or IFIH1), and LGP2 (Laboratory of Genetics and Physiology 2 or DHX58) (68, 10-66 12).

67 RLRs are composed of three different protein domains: a C-terminal domain 68 (CTD) (that mediates initial RNA binding and confers substrate specificity); a central 69 ATPase/DExD/H box RNA helicase domain (which regulates dsRNA binding and ATP 70 hydrolysis; Hel1, Hel2 and Hel2i insertion); and an N-terminal domain with two tandem 71 repeats (its function is the recruitment and activation of caspase; CARD), this last domain 72 is absent in LGP2 (7, 13-15).

73 The interaction of the C-terminal domain of the RIG-I and MDA5 receptors with 74 the pathogenic dsRNA produces a series of changes in the helicase domain that results in 75 the activation of the N-terminal caspase domain of CARDs by interacting with the

76 CARDs of the MAVS adapter (oligomerization) (16) activating NF-kB (transcription
77 factor) and IRF3 and IRF7 (IFN regulatory factors 3 and 7).

78 RIG-I and MDA5 recognize different viral RNA structures. RIG-I detects ssRNA
79 carrying a 5´triphospate residue and short dsRNAs (1kb) (17-19) and MDA5 recognize
80 longer dsRNAs (1-5kb) (18,20).

81 Due to the loss of the CARDs domain, LGP2 is not capable of producing an IFN
82 response but has been shown to exert a negative regulation (interfering self association)
83 on RIG-I (12, 18, 20) and a positive regulation (enhancing MDA5-RNA interactions) on
84 MDA5. (15, 22-28).

85 RIG-I interacts primarily with measles RNA viruses, hepatitis C, vesicular
86 stomatitis, influenza A, respiratory syncytial, Newcastle disease, and rabies (25, 26, 29-
87 32), MDA5 detects picornaviruses (rhinovirus, encephalomyocarditis virus, and
88 Coxsackie B virus), (33-35) while flaviviruses (dengue type 2) or reoviruses have been
89 detected by both RIG-I and MDA5 (30). Furthermore, LGP2 is involved in the generation
90 of the MDA5-triggered IFN response by hepatitis D virus (36).

91 RLR genetic polymorphisms have been involved in viral infection reponses (37,
92 38), inflammatory (39) and autoimmune diseases (40-42) being able to modify the
93 function of these receptors.

94 Another possible mechanism for regulating the function of the corresponding
95 protein is the formation of different specific isoforms so that can further inhibit or activate
96 it. This is produced by alternative splicing mechanisms of the RNA of the protein, as has
97 already been identified in splicing regulation T cells (43).

98 The objective of the present work was to determine the presence of real (not predicted)
99 alternative transcripts of LGP2 in a panel of healthy unrelated individuals, their structure,

100 their possible function, and their evolution with respect to non-human primates, in order
101 to understand the mechanisms involved in RLRs receptors in the innate immune response.
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124

125 **MATERIALS AND METHODS**

126 **Samples**

127 LGP2 mRNA transcripts were analyzed in a total of 15 healthy, anonymous,
128 unrelated blood donors obtained from the Transfusion Center of the Comunidad de
129 Madrid (Madrid, Spain). Peripheral blood mononuclear cells (PBMCs) were obtained
130 from buffy coat bags with EDTA by centrifugation in a Ficoll-Hypaque density gradient
131 (Corning, NY, USA). PBMCs were cryopreserved and stored in liquid nitrogen until the
132 moment of analysis.

133 **Total RNA isolation**

134 Total RNA was extracted from PBMCs using the RNeasy Mini kit (Qiagen Iberia,
135 Madrid, Spain), as stated by the manufacturer's protocols.

136 After lysing the cells and adding ethanol, the lysate was added to a silica
137 membrane removing all possible contaminants. A specific process to eliminate
138 contaminating genomic DNA was carried out with the RNAse-Free DNAse Set Kit and
139 the RNA concentration was measured with the NanoDrop 2000 spectrophotometer
140 (Thermo Scientific, Wilmington, DE, USA).

141

142 **DNA extraction**

143 Whole genomic DNA was extracted from blood with the QIAamp DNA Blood
144 Mini Kit (Qiagen Iberia, Madrid, Spain), following the manufacturer's instructions. After
145 proteinase treatment, the resulting lysate has the conditions to allow optimal DNA
146 binding to the QIAamp silica membrane before loading the sample onto the spin column.
147 DNA bound to the QIAamp membrane is washed away to improve the purity of the eluted
148 DNA. The washing conditions ensured complete removal of any residual contaminants
149 without affecting DNA binding. The purified DNA is finally eluted in concentrated form

150 in water. DNA concentration and purity were measured by NanoDrop 2000
151 spectrophotometer.

152 **RT-PCR amplification**

153 One μg of total RNA was reverse transcribed to cDNA using the NZY first-strand cDNA
154 synthesis RT kit with dual primer, oligo (dT), and random hexamers (Nzytech, Portugal),
155 according to the manufacturer's protocol. Then, samples were treated with NZY RNase
156 H to eliminate the binding of RNA to cDNA. This procedure is mainly recommended
157 when using cDNA in PCR amplification, especially for some targets (>1 kb) that may
158 require RNA-free DNA as a template. Removing the RNA will increase the sensitivity of
159 the PCR step. The obtained cDNA was used directly in PCR or stored at −20 °C. cDNA
160 PCR was performed using NZYTaq II 2x Colorless Master Mix (Nzytech, Portugal),
161 according to the manufacturer's protocol. The primers shown in Table 1 were used for
162 general and specific amplification of LGP2 to determine expression of the alternative
163 transcripts previously described. Amplification was performed at 5 min at 95ºC for initial
164 denaturation, followed by 35 cycles of 94ºC (45 s, denaturation), 60ºC (45 s, annealing)
165 and 72ºC (3 min, elongation), and 30 min at 72ºC for the final elongation. PCR products
166 were electrophoresed on a 1.5% agarose gel stained with ethidium bromide (0.5 mg/ml).
167 NZY DNA Ladder VIII (Nzytech, Portugal) was used as a molecular weight reference.
168 Gel images were captured using the G:BOX F3 gel imaging system (Syngene,
169 Cambridge, UK).

170 **DNA amplification introns by PCR**

171 To obtain intron sequences between exons 1, 2, 3 and 10, 11, specific DNA PCR
172 amplification was performed with the corresponding primers (Table 1) using the same
173 conditions described above for cDNA amplification.

174 **Sequencing of PCR products**

175 The obtained LGP2 PCR products were sequenced using the same specific
176 amplification primers (Table 1). PCR products yielding one band were purified using the
177 EXOSAP-IT® methodology (Life technologies, Thermofisher, Waltham, MA, USA).
178 Extraction and gel purification were performed for each gel band using IlustraTM GFX
179 PCR DNA and Gel Band Purification Kit (GE Healthcare, United Kingdom). The
180 obtained products were sequenced in both directions in ABI PRISM 3730 genetic
181 analyzer (Applied Biosystems) using BigDye terminator cycle sequencing kit (Applied
182 Biosystems, Thermofisher, Waltham, MA, USA). To avoid PCR errors, each sample was
183 sequenced from two different PCR amplifications. The obtained sequences were
184 compared with reference LGP2 cDNA sequences obtained from Genbank (see below).

185 **Sequence analysis and accession numbers of the sequences used**

186 The homology between sequences and the translation of different transcripts was
187 analyzed using Blast© (National Center for Biotechnology Information, Bethesta, MD,
188 USA). The alignment of the obtained sequences was performed with MEGA 11 software
189 (65) with the corresponding published sequences. Neighbor-joining tree was obtained
190 with 500 boostraps from exon 1 to exon 3. Protein structures and configuration of the
191 domains were obtained in all the isoforms based on previously described model (Jiang et
192 al., 2023) with SWISS MODEL software (Computational Structural Biology Group,
193 Institute Swiss Bioinformatics and Biozentrum, University of Basel) (66). Briefly,
194 SWISS-MODEL is a web-based integrated service dedicated to protein structure
195 homology modelling. It guides the user in building protein homology models at different
196 levels of complexity. The Model Results page of SWISS-MODEL provides the user with
197 an essential; first glance view of a homology model showing ligands, global and local
198 quality, target-template alignments. SWISS-MODEL is freely available at
199 https://swissmodel.expasy.org.

Nine alternative transcripts for LGP2 (DHX58) have been described in humans, and eight of them have been predicted according to Genbank: V1: **NM024119.3**: VX1 **XM054317165.1**; VX2: **XM054317166.1**; VX3: **XM054317167.1**; VX4: **XM047436724.1**; VX5: **XM054317164.1**; VX1.1: **XM047436725.1**; VX2.1: **XM047436726.1**; and VX3.1: **XM047436727.1**) (the letter "V" has been added to the names to represent the corresponding variant; in addition, ".1" has been added to differentiate two sequences with the same name but different Genbank accession numbers). In non-human primates, Bonobo (*Pan paniscus*) has 3 predicted transcripts, VX1: **XM008966476.5**; VX2: **XM034941844.2**; and VX3: **XM034941847.2**; Chimpanzee (*Pan troglodytes*) has 6 predicted transcripts VX1: **XM016931789.3**; VX2: **XM054670881.1**; VX3: **XM016931791.3**; VX4: **XM016931790.3**; VX5 **XM016931792.2**; VX6: **XM054670882.1**; and Gorilla (*Gorilla gorilla*) has 3 predicted transcripts, VX1: **XM004041703.3**; VX2: **XM019026949.2**; and VX3: **XM055388114.1**.

225 **RESULTS**

226 **Human LGP2 mRNA transcripts**

227 To perform a comprehensive evaluation of human LGP2 transcripts, including the
228 possibility of novel insertions, deletions, and splicing events, human nuclear cells were
229 obtained, and generic and transcript-specific primers (Table 1) were designed to identify
230 the predicted and novel transcripts. Once new transcripts were defined, a new set of
231 specific primers was designed. One transcript was described as such (named V1; V means
232 "variant") and eight as "predicted" (named VX1-VX5 and VX1.1, VX2.1 and VX3.1; dot
233 and 1 are added to differentiate these sequences from those of VX1, VX2 and VX3 since
234 they have the same name in Genbank although with different accession numbers). VX2
235 and VX2.1 are identical sequences (see Genbank accession numbers).

236 The most important characteristics of LGP2 transcripts were as follows (Figure 1
237 and Supplementary Figure 1): V1 showed 14 exons with a total of 2,591 nucleotides and
238 it was translated into a protein from nucleotides 205 to 2241 with 678 amino acids; VX1
239 had the same exons as V1 and a 309-nucleotides partial insertion belonging to intron 1,
240 which would produce the same protein; VX2 was similar to V1 lacking exons 12 and 14
241 with a different protein structure; VX3 showed an insertion of 399 nucleotides belonging
242 to the entire sequence of intron 1 and lacked exons 9, 11, 12, 13, and 14, yielding a
243 different protein; VX4 was similar to V1 with the insertion of 2 nucleotides from intron
244 1 that produced the same protein. VX1.1, VX3.1, and VX5, showed an insertion before
245 exon 1 of 9, 11, and 16 nucleotides, respectively, and lacked the first nucleotide of exon
246 1 (Supplementary Figure 1). Additionally, VX5 shared 398 of 399 insertion nucleotides
247 of intron 1 in VX3, but including all exons. VX1.1 had the same structure as VX1
248 (insertion of 309 nucleotides and all exons), while VX3.1 had the same structure as VX5
249 (insertion of 398 nucleotides of intron 1 and also deletion of exons 9, 11-14) with the

250 above insert/delete exception mentioned. The only previously described V1 transcript
251 was found in all samples. Regarding the previously predicted transcripts, VX1, VX2
252 (same sequence of VX2.1), and VX4 were also found. The rest of the predicted transcripts
253 VX3, VX5, VX1.1, and VX3.1 were not found in any of the analyzed samples.

254 On the other hand, eight new LGP2 transcripts were found in all samples: V2 had
255 a similar structure to V1 but with the deletion of exon 2 and an insertion of 192
256 nucleotides belonging to intron 2; V3 showed the same structures as V2 but with a partial
257 insertion of 58 nucleotides of intron 3 and subsequent duplication of exon 3; V4 showed
258 similar structure to Vx1 including the insertion of 309 nucleotides of intron 1 and 219
259 nucleotides of intron 2; V5 had the same structure as V4 but with longer insertion of 399
260 nucleotide of intron 1 that was common to VX3.1, VX5, and VX3; V6 had a similar
261 structure to V1 with the same 219 nucleotide insertion of intron 2 as observed in V4 and
262 V5. In addition, V6 had an insertion of 90 nucleotides belonging to intron 10; V7 showed
263 the same structures as V6 but with insertion of short intron 2 of 192 nucleotides and
264 without insertion of intron 10; V8 had the same structure as V4 but with a short insertion
265 of 192 nucleotides; and V9 had the same structure as V5 but with a short insertion of 192
266 nucleotides.

267 Nucleotide sequences of the cDNA variants obtained were submitted to the
268 GenBank and the following accession numbers were assigned: Hosa-V2: **OR199842**,
269 Hosa-V3: **OR965977**, Hosa-V4: **OR778083**, Hosa-V5: **OR778084**, Hosa-V6:
270 **OR778085**, Hosa V7: **PP350412**, Hosa-V8: **PP350413**, and Hosa-V9: **PP350414**. To
271 determine if intron insertions (1, 2, 3, 10) were partial or total, a fragment of 1407
272 nucleotides from genomic DNA including the partial UTR region 5', exon 1, intron 1,
273 exon2, intron 2, exon 3, intron 3, and partial exon 4 were also sequenced (Supplementary

274 Figure 1, Genbank accession number: **PP350411**). The DNA sequence of intron 10 was
275 identical to Hosa-V6 transcript (Genbank accession number: **OR778085**).

276 **LGP2 mRNA transcripts in non-human primates and comparison with those in**
277 **humans**

278 To determine whether the insertions/deletions of different LGP2 transcripts were
279 generated in humans or, previously, in closest ancestors, the sequences of these transcripts
280 were also analyzed in non-human primates such as chimpanzee (*Pan troglodytes*, Patr),
281 bonobo (*Pan paniscus*, Papa), and gorilla (*Gorilla gorilla*, Gogo). A total of 12 predicted
282 non-human primate LGP2 transcripts (6 Patr, 3 Papa and 3 Gogo) were described in
283 Genbank (see Materials and Methods) and used for comparison with those from humans.

284 *Pan troglodytes LGP2 mRNA transcripts*

285 Patr-VX1 showed a structure similar to that of human and specifically, 392
286 nucleotides of intron 1 were inserted with a 96% homologous sequence (Table 2) to
287 human Hosa-VX3, Hosa-VX5, and Hosa-VX3.1; in addition, Patr-VX1 had an insertion
288 of 151 nucleotides before exon 1 (a sequence of 11 nucleotides appeared in human Hosa-
289 VX5, Hosa-VX3.1, and Hosa-VX1.1), and other in intron 3 of 227 nucleotides. Part of
290 the latter appeared in human Hosa-V3 with 81% homology. Patr-VX2 had the same
291 structure as Patr-VX1 but without the insertion of intron 3; Patr-VX3 was identical to
292 Patr-VX2 but lacked the intron 1 insertion; Patr-VX4 had the same sequence as Patr-VX2
293 but with the shorter insertion of 309 nucleotides similar to that of human Hosa-VX1,
294 Hosa-VX1.1, Hosa-V4 and Hosa-V8 with 96% homology; Patr-VX5 showed the same
295 sequence as Patr-VX2 but with a 192-nucleotide intron 2 insertion similar to that of
296 human Hosa-V2, Hosa-V3, Hosa-V7, Hosa-V8, and Hosa-V9 (98% homology), as well
297 as the 227 nucleotide insertion of intron 3 as in Patr-VX1 and VX6; Patr-VX6 was the

298 same sequence as Patr-VX1 but with the deletion of exons 12 and 14 (as in Hosa-VX2
299 sequence) (Figure 1, Supplementary Figure 1, and Table 2).

300 *Pan paniscus LGP2 mRNA transcripts*

301 Papa-VX1 showed the same 14-exon structure of chimpanzees and humans with
302 the insertion of 307 nucleotides of intron 1 (similar to Patr-VX4, 99% homology) and
303 Hosa-V4, Hosa-V8, Hosa-VX1.1, and Hosa-VX1 (96% homology), as well as an
304 insertion of 3 nucleotides of the sequence before the start of exon 1; Papa-VX2 had a
305 similar structure to Hosa-VX4 with the insertion of the same first 2 nucleotides of intron
306 1 and, in particular, a deletion of 5 nucleotides from exon 1; and Papa-VX3 had a similar
307 structure to Patr-VX1, Patr-VX2, Patr-VX5, Patr-VX6 (99/100% homology) and Hosa-
308 V5, Hosa-V9, Hosa-VX5, Hosa-VX3.1, Hosa-VX3. (96 % homology), regarding to the
309 insertion of 396 nucleotides of intron 1, but with an insertion of 4 nucleotides in the
310 sequence before exon 1. In addition, exon 12 and the first 21 nucleotides of exon 13 were
311 deleted (Figure 1, Supplementary Figure 1, and Table 2).

312 *Gorilla gorilla LGP2 mRNA transcripts*

313 Gogo-VX1 has a similar structure to Hosa-V1, without insertions and with the
314 deletion of exon 1. Gogo-VX2 had a similar structure to Hosa-V2 with the deletion of
315 exon 2 and the insertion of 192 nucleotides of intron 2 as in Patr-VX5 (98% homology),
316 Hosa-V2, Hosa-V3, Hosa-V7, Hosa-V8 and Hosa-V9 (99% homology). Only the last 29
317 nucleotides of exon 1 were transcribed; and Gogo-Vx3 was similar to Gogo-VX1 but with
318 the same deletion of exon 2 and a 57-nucleotide insertion of intron 1 (Figure 1,
319 Supplementary Figure 1, and Table 2).

320 **Phylogenetic tree analysis and LGP2 evolution**

321 A neighbor-joining tree was constructed using sequences from humans and non-
322 human primates, including those from exon 1 to exon 3. With this phylogenetic analysis,

323 the different genetic structures of the mRNAs of the species considered could be
324 compared. The phylogenetic tree had three well-differentiated groups that showed the
325 evolution from gorillas to humans. In gorillas there were no LGP2 transcripts with a
326 partial 3'UTR or intron 1. These transcripts appeared in chimpanzees and bonobos,
327 indicating that these transcripts may have been generated from gorillas and others directly
328 from chimpanzees/bonobos. These groups would be established as evolutionary lineages
329 as follows (Figure 1): Lineage A showed two differentiated subgroups A1 and A2. A1
330 subgroup would include Patr-VX2, Patr-VX6, Patr-VX1, and Patr-VX4 sequences. The
331 first three had the same sequence of intron 1 and the fourth shared a shorter sequence than
332 the previous ones but with the same sequence. Papa-VX3 and Papa-VX1 sequences with
333 an intron 1 sequence similar to the previous ones (99% homology) were also included.
334 A2 subgroup would include human Hosa-VX1 and Hosa-VX1.1 sequences with a shorter
335 intron 1 insertion sequence (309 nucleotides), and Hosa-VX3, Hosa-VX5, Hosa-VX3.1
336 with a complete sequence of intron 1 (399 nucleotides) (99% homology between them
337 and 96% with respect to group A1). Lineage B included Gogo-VX2, Hosa-V2, Hosa-V3,
338 Hosa-V6, and Hosa-V7 sequences. All were characterized by having an intron 2 insertion
339 of 219 nucleotides in Hosa-V6 or 192 nucleotides in Gogo-VX2 (99% homology), Hosa-
340 V2, Hosa-V3, and Hosa-V7 (100% homology). This group also included Patr-VX5 (98%
341 homology), Hosa-V4, Hosa-V5, Hosa-V8, and Hosa-V9 which, in addition to including
342 the shortest intron 2 (Patr-VX5, Hosa -V8, and Hosa-V9) or complete intron 2 insertion
343 (Hosa-V4 and Hosa-V5), also had the shorter intron 1 insertion (Hosa-V4 and Hosa-V8)
344 or complete intron 1 insertion (Patr-VX5 , Hosa-V5, and Hosa-V9). This insertion
345 appeared in lineage A, so this group could be called lineage A/B. Lineage C grouped
346 Gogo-VX1, Gogo-Vx3, Papa-VX2, Patr-VX3, Hosa-V1, Hosa-VX2, Hosa-VX2.1, and
347 Hosa-VX4 sequences lacking intron 1 or intron insertions 2. There was only a partial 2-

348 nucleotide intron 1 insertion in Hosa-VX4 and Papa-VX2, and a 3'UT insertion sequence
349 in Patr-VX3 like Patr group A1.
350 From the sequences obtained we determined that gorillas would have lineages B
351 and C, bonobos would have lineages A and C, chimpanzees would have lineages A, C
352 and a hybrid A/B, and humans would have lineages A, B, C and A/B. We also established
353 that lineage B and C came from gorillas, lineage A from chimpazee/bonobo, and hybrids
354 A/B seemed to come from chimpanzee. These sequences had a high degree of homology
355 and maintained the 3 lineages, so there was an evolutionary pressure to maintain them for
356 more than 10 million years in evolution, because they probably gave an evolutionary
357 advantage to the innate immune response.
358 **Protein structure of the human isoforms**
359 From the transcripts obtained, the presence of four isoforms can be observed (Figure 3).
360 One complete isoform (IS1) would include transcripts V1, V2 (loss of exon 2 and
361 insertion of partial sequence of intron 2), V3 (insertion of partial sequence of intron3 and
362 duplication of exon 3), V4 and V5 (with insertions of partial and complete sequence of
363 intron 1, respectively, and insertion of complete sequence of intron 2), V7 (with insertion
364 of partial sequence of intron 2), V8, V9 (with insertion of the shortest or complete
365 sequence of intron 1 and the shortest insertion of intron 2), VX1 and VX1.1 (with
366 insertion of the partial sequence of intron 1), VX4 (with insertion of 2 nucleotides of
367 intron 1), and VX5 (with insertion of the complete sequence of intron 1), producing all of
368 them functional proteins. The second isoform (IS2), which would include transcripts VX2
369 and VX2.1, would lose the C-terminal domain and a 19 amino acid fragment of the pincer.
370 The third isoform (IS3) would include V6 transcript and would lose the C-terminal region,
371 the pincer, and 48 amino acids from Hel2 domain. And the fourth isoform (IS4) would

372 lose the C-terminal region, the pincer, and the entire Hel2 domain, including VX3 and

373 VX3. 1 transcripts. In these cases, the functionality of the LGP2 would be altered.

374 **DISCUSSION**

375 Two types of RNA splicing have been defined, the constitutive one which consists
376 of the deletion of the introns of a gene for the creation of the pre-RNA and the alternative
377 (AS) one in which there is deletion of exons and maintenance of introns (MI), giving rise
378 to different RNAs of the same gene and therefore the different corresponding proteins.
379 These mechanisms contribute to the regulation of the expression of a given protein either
380 to activate it, to inhibit it or to give rise to a different function (44, 45).

381 Until now, the maintenance of introns in alternative splicing mechanisms has not
382 been considered important or has not even been taken into account in humans and other
383 mammals. (46-48).

384 In this study, we observed that 14 human transcripts contained insertions of
385 introns 1 and 2 (5 with complete and 5 partial insertion of intron 1; 3 with total insertion
386 and 5 with partial insertion of intron 2; 1 with complete insertion of intron 10) and 7
387 transcripts contained exon deletions. The insertion of introns in the transcripts is already
388 present in 11 of non-human primates such as Gorilla, Bonobo, and Chimpanzee.
389 Furthermore, there was total or partial deletion of exons in 6 non-human primates.

390 MI is often associated with gene expression downregulation through a destruction
391 of non-functional RNA mechanism that does not produce a protein or the one that is
392 generated is truncated as when exons are lost (see Figures 1 and 4) (49).

393 The possibility that these AS mechanisms regulate the immune response is
394 currently being raised. In fact, several key proteins have recently been identified in
395 splicing regulation of T cells (43). We demonstrated that LGP2 can be added to genes
396 that are regulated by alternative splicing belonging to one of the RLRs, thereby
397 contributing to the innate immune response.

398 The immune and reproductive systems present genes in which positive selection
399 has been demonstrated in the different species of non-human primates (50-54).
400 These results are consistent, since the survival to certain infectious pathogens
401 (different immune system) and reproductive competence would explain this positive
402 selection of the corresponding genes.
403 The evolutionary separation between humans, gorillas and chimpanzees is around
404 10 million years for gorillas/humans and between 6.5 and 7.5 million years for
405 chimpanzees/humans.
406 In this study, we defined an evolutionary process of positive selection for the
407 alternative transcripts found for LGP2 in the studied species (Figure 2). The inheritance
408 of the 3 lineages obtained was observed, two from the gorilla (lineages B and C) and one
409 (lineage A) from chimpanzees/bonobos but not found in gorillas. Humans have also
410 inherited a hybrid lineage (lineage A/B) from the chimpanzee and that has not been found
411 yet in bonobos or gorillas. In the different lineages, intron 1 has been inserted totally or
412 partially in lineage A and intron 2 partially or totally in lineage B, therefore the insertion
413 of introns was confirmed as one of the most stable evolutionary mechanisms for AS.
414 From our data, it has been possible to verify the positive selection of alternative
415 transcripts from the gorilla through chimpanzees and bonobos to humans related to the
416 innate immune system, in this case the LGP2 receptor, as had already been described
417 previously for genes of the adaptive immune system (43). In this case, the homology
418 between the different lineages described is 96-99%, which means that this system has
419 been selected for many millions of years, probably due to its effectiveness in defending
420 the organism against the recognition of specific viruses by RLR receptors.
421 The CARDs domain in RIG-I and MDA5 is conformationally in a non-active
422 structure (19). When the infectious process occurs, the viral RNAs bind mainly to the

423 CTD and helicase domains, starting the activity of the CARDs domain, then this domain
424 interacts with the MAVS protein to produce an antiviral response (55, 56). LGP2 lacks
425 CARD domains, which prevents it from recognizing RNA (21, 57, 58) and, therefore,
426 with different functionality.

427 LGP2 has a double function with respect to RIG I and MDA5. Regarding RIG-I,
428 it has an inhibitory function in three different ways, the first consists of the union of the
429 C-terminal domain of LGP2 and that of RIG-I so that it cannot be recognized the viral
430 RNA (21, 59), the second consists of blocking the interaction between RIG-I and MAVS
431 preventing this activation pathway (60).and thirdly it also inhibits the activation of IRF3,
432 NFkB and the cGAMP/STING pathways (61-64).

433 In the case of MDA5, LGP2 may have an activating or inhibitory function. In vitro
434 experiments have shown that increasing the presence of LGP2 leads to an improvement
435 in the interaction between MDA5 and the viral dsRNA. This happens at low
436 concentrations of LGP2, if we increase its concentration to high levels the opposite effect
437 occurs, producing an inhibition effect (14).

438 In three of the four isoforms described, either the CTD domain or this domain and
439 the one corresponding to the DExD/H RNA helicase Hel2 box are partially or totally lost,
440 becoming truncated isoforms that prevent the function of LGP2 and therefore, becoming
441 a regulation mechanism at the level of the protein formed.

442 In conclusion, this is the first study that detailed the extensive genomic
443 organization and complex transcriptional regulation of human LGP2. Eight new
444 transcripts, one previously described, and four predicted transcripts were identified for
445 LGP2, demonstrating considerably greater diversity in human LGP2 transcripts than
446 previously thought. This diversity arises from AS, and these patterns obtained would be
447 in a transspecies evolutionary inheritance mode. We can observe that there are two

448 alternative splicing mechanisms for the regulation of LGP2, one characterized by the
449 creation of multiple pre-mRNAs by insertion of introns without protein variation, and
450 other by the loss of exons, which would give rise to truncated proteins. These would be
451 the mechanisms of regulation of LGP2 expression and its function as an inhibitor or
452 enhancer of the functions of the other RLRs, RIG-I and MDA5.

453
454
455
456
457
458
459
460
461
462
463
464
465 **ACKNOWLEDGEMENTS**

466 We greatly appreciate all the participants for their contribution to this study.

467 **FUNDING**

468 This work was supported grant PI22CIII/00059 funded by the Strategic Action in
469 Health of the Instituto de Salud Carlos III (ISCIII) and CIBERINFEC, co-financed by the
470 European Regional Development Fund (ERDF) "A way to make Europe". The work of
471 Guiomar Casado is funded by the Consejería de Educación, Universidades, Ciencia y
472 Portavocía of the Comunidad de Madrid (Spain). The work of Clara Sánchez-Menéndez
473 is financed by Programa Investigo, FIBio HRC-IRYCIS, co-financed by FEDER. The
474 work of Montserrat Torres is financed by CIBERINFEC (CB21/13/00015).

475 **CONFLICTS OF INTEREST**

476 The authors declare no conflict of interests.

477 **AUTHOR CONTRIBUTIONS**

Jorge Martinez-Laso: Conceptualization, Methodology, Formal analysis, Writing-Original Draft, Supervision, Project administration. Isabel Cervera: Validation, Formal analysis, Investigation. Marina S. Martinez-Carrasco: Formal analysis, Investigation, Software. Veronica Briz: Investigation, Formal analysis, Resources. Celia Crespo-Bermejo: Validation, Investigation. Clara Sánchez-Menéndez: Formal analysis, Investigation, Software. Guiomar Casado-Fernández: Validation, Investigation, Software. Montserrat Torres: Formal analysis, Investigation, Writing-Review and Editing.

Mayte Coiras: Conceptualization, Formal analysis, Resources, Writing-Review and Editing, Supervision, Funding acquisition, Project administration.

REFERENCES

1. Akira, S., Uematsu, S. and Takeuchi, O. (2006) Pathogen recognition and innate immunity. Cell, **124**, 783–801.

2. Beutler, B., Eidenschenk, C., Crozat, K., Imler, J. L., Takeuchi, O., Hoffmann, J. A. and Akira, S. (2007) Genetic analysis of resistance to viral infection. Nat. Rev. Immunol., **7**, 753–766.

3. Meylan, E., Tschopp, J. and Karin, M. (2006) Intracellular pattern recognition receptors in the host response. Nature, **442**, 39–44.

4. Duan, T., Du, Y., Xing, C., Wang, H.Y. and Wang, R.F. (2022) Toll-Like Receptor Signaling and Its Role in Cell-Mediated Immunity. Front. Immunol., 10.3389/fimmu.2022.812774.

5. Wicherska-Pawłowska, K., Wróbel, T. and Rybka, J. (2021) Toll-Like Receptors (TLRs), NOD-Like Receptors (NLRs), and RIG-I-Like Receptors (RLRs) in Innate

503 Immunity. TLRs, NLRs, and RLRs Ligands as Immunotherapeutic Agents for
504 Hematopoietic Diseases. Int. J. Mol. Sci., 10.3390/ijms222413397.
505 6. Loo, Y.M. and Gale Jr. M. (2011) Immune Signaling by RIG-I-like Receptors.
506 Immunity, 34, 680-692.
507 7. Baldaccini, M. and Pfeffer, S. (2021) Untangling the roles of RNA helicases in antiviral
508 innate immunity. PLoS Pathog. 10.1371/journal.ppat.1010072.
509 8. Goubau, D., Deddouche, S. and Reis e Sousa, X. (2013) Cytosolic sensing of viruses.
510 Immunity, **38**, 855 – 869.
511 9. Schneider, W. M., Chevillotte, M.D. and Rice, C. M. (2014) Interferon-stimulated
512 genes: a complex web of host defenses. Annu. Rev. Immunol., **32**, 513 – 545.
513 10. Wu, J. and Chen, Z. J. (2014) Innate immune sensing and signaling of cytosolic
514 nucleic acids. Annu. Rev. Immunol., **32**, 461-488.
515 11. Schlee, M. and Hartmann, G. (2016) Discriminating self from non-self in nucleic acid
516 sensing. Nat. Rev. Immunol., **16**, 566 – 580.
517 12. Vasseur, E., Patin, E., Laval, G., Pajon, S., Fornarino, S., Crouau-Roy, B. and
518 Quintana-Murci, L. (2011) The selective footprints of viral pressures at the human RIG-
519 I-like receptor family. Hum. Mol. Genet., **20**, 4462-4474.
520 13 Rehwinkel, J. and Gack, M.U. (2020) RIG-I-like receptors: their regulation and roles
521 in RNA sensing. Nat. Rev. Immunol., 20, 537–551.
522 14. Bruns, A. M. and Horvath, C. M. (2015) LGP2 synergy with MDA5 in RLR-mediated
523 RNA recognition and antiviral signaling. Cytokine, **74**, 198–206.
524 15. Lassig, C. and Hopfner, K. P. (2017) Discrimination of cytosolic self and non-self
525 RNA by RIG-I-like receptors. J. Biol. Chem., **292**, 9000–9009.
526 16. Rout, A. K., Udgata, S. R., Dehury, B., Pradhan, S. P., Swain, H.S., Behera, B. K. and
527 Das, B.K. (2019) Structural bioinformatics insights into the CARD-CARD interaction

528 mediated by the mitochondrial antiviral-signaling protein of black carp. J. Cell. Biochem.,
529 120, 12534-12543.
530 17. Schlee, M., Roth, A., Hornung, V., Hagmann, C.A., Wimmenauer, V., Barchet, W.,
531 Coch, C., Janke, M., Mihailovic, A., Wardle, G., Juranek, S., Kato, H., Kawai, T., Poeck,
532 H., Fitzgerald, K. A., Takeuchi, O., Akira, S., Tuschl, T., Latz, E., Janos Ludwig, J. and
533 Hartmann, G. (2009) Recognition of 5′ triphosphate by RIG-I helicase requires short blunt
534 double-stranded RNA as contained in panhandle of negative-strand virus. Immunity, **31**,
535 25–34.
536 18. Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K.,
537 Hiiragi, A., Dermody, T.S., Fujita, T. and Akira, S. (2008) Length-dependent recognition
538 of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma
539 differentiation-associated gene 5. J. Exp. Med., **205**, 1601–1610.
540 19. Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale,
541 M. Jr, Inagaki, F. and Fujita, T. (2008) Nonself RNA-sensing mechanism of RIG-I
542 helicase and activation of antiviral immune responses. Mol. Cell, **29**, 428–440.
543 20. Nallagatla, S.R., Toroney, R. and Bevilacqua, P.C. (2008) A brilliant disguise for self
544 RNA: 5′-end and internal modifications of primary transcripts suppress elements of innate
545 immunity. RNA Biol., **5**, 140–144.
546 21. Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura,
547 T., Fujita, T., Akira, S. and Takeuchi, O. (2010) LGP2 is a positive regulator of RIG-I-
548 and MDA5-mediated antiviral responses. Proc. Natl. Acad. Sci. U.S.A., 107, 1512-7151.
549 22. Bruns, A. M., Leser, G.P., Lamb, R.A. and Horvath, C. M. (2014) The innate immune
550 sensor LGP2 activates antiviral signaling by regulating MDA5-RNA interaction and
551 filament assembly. Mol. Cell., **55**, 771–781.

552 23. Venkataraman, T., Valdes, M., Elsby, R., Kakuta, S., Caceres, G., Saijo, S., Iwakura,
553 Y. and Barber, G.N. (2007) Loss of DExD/H box RNA helicase LGP2 manifests disparate
554 antiviral responses. J. Immunol., **178**, 6444–6455.
555 24. Hwang, S.Y., Sun, H.Y., Lee, K.H., Oh, B.H., Cha, Y.J., Kim, B.H., Yoo, J.Y. (2012)
556 5′-Triphosphate-RNA-independent activation of RIG-I via RNA aptamer with enhanced
557 antiviral activity. Nucleic Acids Res., 40, 2724–2733.
558 25. Brisse, M. and Ly, H. (2019) Comparative Structure and Function Analysis of the
559 RIG-I-Like Receptors: RIG-I and MDA5. Front. Immunol., 10.3389/fimmu.2019.01586.
560 26. Kell, A.M. and Gale, M. Jr. (2015) RIG-I in RNA virus recognition. Virology, 479,
561 110-121.
562 27. Duic, I., Tadakuma, H., Harada, Y., Yamaue, R., Deguchi, K., Suzuki, Y., Yoshimura,
563 S.H., Kato, H., Takeyasu, K. and Fujita, T. (2020) Viral RNA recognition by LGP2 and
564 MDA5, and activation of signaling through step-by-step conformational changes. Nucleic
565 Acids Res., 48, 11664-11674.
566 28. Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A. J., Zillinger, T., Goldeck, M.,
567 Schuberth, C., van der Veen, A. G., Fujimura, T., Rehwinkel, J., et al. (2014) Antiviral
568 immunity via RIG-I-mediated recognition of RNA bearing 5´-diphosphates. Nature, **514**,
569 372 – 375.
570 29. Ouyang, Y., Liao, H., Hu, Y., Luo, K., Hu, S. and Zhu, H. (2022) Innate Immune
571 Evasion by Human Respiratory Syncytial Virus. Front. Microbiol.
572 10.3389/fmicb.2022.865592.
573 30. Loo, Y. M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L.,
574 Akira, S., Gill, M. A., Garcia-Sastre, A., Katze, M. G. et al. (2008) Distinct RIG-I and
575 MDA5 signaling by RNA viruses in innate immunity. J. Virol., **82**, 335–345.

576 31. Wang, C., Wang, T., Duan, L., Chen, H., Hu, R., Wang, X., Jia, Y., Chu, Z., Liu, H.,
577 Wang, X. et al. (2022) Evasion of Host Antiviral Innate Immunity by Paramyxovirus
578 Accessory Proteins. Front. Microbiol., 10.3389/fmicb.2021.790191. 32. Mikkelsen, S.
579 S., Jensen, S. B., Chiliveru, S., Melchjorsen, J., Julkunen, I., Gaestel, M., Arthur, J. S.,
580 Flavell, R. A., Ghosh, S. and Paludan, S.R. (2009) RIG-I-mediated activation of p38
581 MAPK is essential for viral induction of interferon and activation of dendritic cells:
582 dependence on TRAF2 and TAK1. J. Biol. Chem., **284**, 10774–10782.
583 33. Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R.A., Diamond,
584 M.S. and Colonna, M. (2006) Essential role of MDA-5 in type I IFN responses to
585 polyriboinosinic: polyribocytidylic acid and encephalomyocarditis picornavirus. Proc.
586 Natl Acad. Sci. USA, **103**, 8459–8464.
587 34. Wang, Q., Nagarkar, D.R., Bowman, E.R., Schneider, D., Gosangi, B., Lei, J., Zhao,
588 Y., McHenry, C.L., Burgens, R.V., Miller, D.J., et al. (2009) Role of double-stranded
589 RNA pattern recognition receptors in rhinovirus-induced airway epithelial cell responses.
590 J. Immunol., **183**, 6989–6997.
591 35. Wang, J. P., Cerny, A., Asher, D. R., Kurt-Jones, E. A., Bronson, R. T. and Finberg,
592 R.W. (2010) MDA5 and MAVS mediate type I interferón responses to coxsackie B virus.
593 J. Virol., **84**, 254–260.
594 36. Gillich, N., Zhang, Z., Binder, M., Urban, S. and Bartenschlager, R. (2023) Effect of
595 variants in LGP2 on MDA5-mediated activation of interferon response and suppression
596 of hepatitis D virus replication. J. Hepatol., **78**, 78–89.
597 37. Hu, J., Nistal-Villan, E., Voho, A., Ganee, A., Kumar, M., Ding, Y., Garcia-Sastre,
598 A. and Wetmur, J.G. (2010) A common polymorphism in the caspase recruitment domain
599 of RIG-I modifies the innate immune response of human dendritic cells. J. Immunol.,
600 **185**, 424–432.

601 38. Pothlichet, J., Burtey, A., Kubarenko, A.V., Caignard, G., Solhonne, B., Tangy, F.,
602 Ben-Ali, M., Quintana-Murci, L., Heinzmann, A., Chiche, J.D., et al. (2009) Study of
603 human RIG-I polymorphisms identifies two variants with an opposite impact on the
604 antiviral immune response. PLoS One, 10.1371/journal.pone.0007582.
605 39. Barreiro, L. B. and Quintana-Murci, L. (2010) From evolutionary genetics to human
606 immunology: how selection shapes host defence genes. Nat. Rev. Genet.**, 11**, 17–30.
607 40. Jermendy, A., Szatmari, I., Laine, A.P., Lukacs, K., Horvath, K.H., Korner, A.,
608 Madacsy, L., Veijola, R., Simell, O., Knip, M. J., et al. (2010) The interferon-induced
609 helicase IFIH1 Ala946Thr polymorphism is associated with type 1 diabetes in both the
610 high-incidence Finnish and the medium-incidence Hungarian populations. Diabetologia,
611 **53**, 98–102.
612 41. Kato, H. and Fujita, T. (2015) RIG-I-like receptors and autoimmune diseases. Curr.
613 Opin. Immunol., 37, 40-45.
614 42. Ferreira, R. C., Pan-Hammarstrom, Q., Graham, R. R., Gateva, V., Fontan, G., Lee,
615 A. T., Ortmann, W., Urcelay, E., Fernandez-Arquero, M., Nunez, C., et al. (2010)
616 Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency.
617 Nat. Genet., **42**, 777–780.
618 43. Martinez, N. M. and Lynch, K. W. (2013) Control of alternative splicing in immune
619 responses: many regulators, many predictions, much still to learn. Immunological
620 Reviews, **253**, 216–236
621 44. Ghigna, C., Valacca, C. and Biamonti, G. (2008) Alternative Splicing and Tumor
622 Progression. Curr. Genom., **9**, 556–570.
623 45. Baralle, F. E. and Giudice, J. (2007) Alternative splicing as a regulator of development
624 and tissue identity. Nat. Rev. Mol. Cell Biol., **18**, 437–451.

625 46. EL-Seedy, A., Pellerin, L., Page, G., and Ladeveze, V. (2023) Identification of Intron
626 Retention in the Slc16a3 Gene Transcript Encoding the Transporter MCT4 in the Brain
627 of Aged and Alzheimer-Disease Model (APPswePS1dE9) Mice. Genes, **14**, 1949.
628 47. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by
629 alternative splicing. Nature, **463**, 457–463.
630 48. Vaquero-Garcia, J., Barrera, A., Gazzara, M. R., Gonzales-Vallinas, J., Lahens, N. F.,
631 Hogenesch, J. B., Lynch, K. W. and Barash, Y. (2016) A new view of transcriptome
632 complexity and regulation through the lens of local splicing variations. eLife,
633 10.7554/eLife.11752.
634 49. Marquez, Y., Hopfler, M., Ayatollahi, Z., Barta, A. and Kalyna, M. (2015)
635 Unmasking alternative splicing inside protein-coding exons definesexitrons and their role
636 in proteome plasticity. Genome Res., **25**, 995–1007.
637 50. Nakamura, T., Fujiwara, K., Saitou, M. and Tsukiyama T. (2021) Non-human
638 primates as a model for human development. Stem Cell Reports, 16, 1093–1103.
639 51. Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J.,
640 Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012) Insights into
641 hominid evolution from the gorilla genome sequence. Nature, **483**, 169–175.
642 52. Locke, D. P., Hillier, L. W., Warren, W. C., Worley, K. C., Nazareth, L. V., Muzny,
643 D. M., Yang, S.P., Wang, Z., Chinwalla, A.T., Minx, P., et al. (2011) Comparative and
644 demographic analysis of orang-utan genomes. Nature, **469**, 529–533.
645 53. George, R. D., McVicker, G., Diederich, Ng. S. B., MacKenzie, A. P., Swanson, W.J.,
646 Shendure, J. and Thomas, J.H. (2011) Trans genomic capture and sequencing of primate
647 exomes reveals new targets of positive selection. Genome Res., **21**, 1686–1694.
648 54. Rogers, J. and Gibbs, R. A. (2014) Comparative primate genomics: emerging patterns
649 of genomecontent and dynamics. Nat. Rev. Genet., **15**, 347–359.

650 55. Kowalinski, E., Lunardi, T., McCarthy, A. A., Louber, J., Brunel, J., Grigorov B,
651 Gerlier, D. and Cusack, S. (2011) Structural basis for the activation of innate immune
652 pattern-recognition receptor RIG-I by viral RNA. Cell, **147**, 423–435.
653 56. Luo, D., Ding, S. C., Vela, A., Kohlway, A., Lindenbach, B. D. and Pyle, A. M. (2011)
654 Structural insights into RNA recognition by RIG-I. Cell, **147**, 409–422.
655 57. Stok J. E., Oosenbrug T., Ter Haar L. R., Gravekamp D., Bromley C. P., Zelenay S.,
656 e Sousa C. R. and van der Veen A. G. (2022) RNA sensing via the RIG-I-like receptor
657 LGP2 is essential for the induction of a type I IFN response in ADAR1 deficiency. EMBO
658 J., 10.15252/embj.2021109760.
659 58. Takahashi, T., Nakano, Y., Onomoto, K., Yoneyama, M. and Ui-Tei, K. (2018) Virus
660 Sensor RIG-I Represses RNA Interference by Interacting with TRBP through LGP2 in
661 Mammalian Cells. Genes, 10.3390/genes9100511.
662 59. Jain, A., Mittal, S., Tripathi, L. P., Nussinov, R. and Ahmad, S. (2022) Host-pathogen
663 protein-nucleic acid interactions: A comprehensive review. Comput. Struct. Biotechnol.
664 J., **20**, 4415–4436.
665 60. Chang, M. X. (2021) The negative regulation of retinoic acid-inducible gene I (RIG-
666 I)-like receptors (RLRs) signaling pathway in fish. Dev. Comp. Immunol.,
667 10.1016/j.dci.2021.104038.
668 61. Malur M., Gale M. and Krug R.M. (2012) LGP2 downregulates interferon production
669 during infection with seasonal human influenza A viruses that activate interferon
670 regulatory factor 3. J. Virol., **86**, 10733–10738.
671 62. Lenoir, J.J., Parisien, J.P. and Horvath, C. M. (2021) Immune regulator LGP2 targets
672 Ubc13/UBE2N to mediate widespread interference with K63 polyubiquitination and NF-
673 kappaB activation. Cell Rep., 10.1016/j.celrep.2021.110175.

674 63. Parisien, J. P., Lenoir, J. J., Mandhana, R., Rodriguez, K. R., Qian, K., Bruns, A.M.
675 and Horvath, C. M. (2018) RNA sensor LGP2 inhibits TRAF ubiquitin ligase to
676 negatively regulate innate immune signaling. EMBO Rep., 10.15252/embr.201745176.
677 64. Quicke, K. M., Kim, K.Y., Horvath, C.M. and Suthar, M.S. (2019) RNA helicase
678 LGP2 negatively regulates RIG-I signaling by preventing TRIM25-mediated caspase
679 activation and recruitment domain ubiquitination. J. Interferon Cytokine Res., **39**, 669-
680 683.
681 65. Tamura, K., Stecher, G. and Kumar, S. (2021) MEGA11: Molecular Evolutionary
682 Genetics Analysis version 11. Mol. Biol. Evol., **38**, 3022-3027.
683 66. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R.,
684 Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018) SWISS-MODEL:
685 homology modelling of protein structures and complexes. Nucleic Acids Res., **46** (W1),
686 W296-W303.

LEGEND TO FIGURES

Figure 1. Genetic structure of the different LGP2 RNA transcripts ordered according to the phylogenetic neighbor-joining tree obtained from exon 1 to exon 3 with 500 boostraps (bootstrap over 50% is shown) is shown. Grey boxes are exon sequences transcribed, White boxes are sequences deleted from exons or introns transcribed, boxes with inclined lines represents the introns sequences transcribed. Hosa=*Homo sapiens*; Papa=*Pan paniscus;* Patr=*Pan troglodytes*; Gogo=*Gorilla gorilla*; V1 represents mRNA transcript 1 real; VX1 represents mRNA transcript predicted as stated in Genbank information. A, B, C represents the postulated lineages of evolution.

Figure 2. Hypothetical generation of the LGP2 RNA transcripts from Gorilla to humans. Grey boxes are exon sequences transcribed, White boxes are sequences deleted from exons or introns transcribed, boxes with inclined lines represents the introns sequences transcribed.Lineages A, B, and C represents the postulated lineages of evolution and are compared between the different species.

Figure 3. Schematic structure of the exon/intron distribution of the LGP2 gene and the molecule resulting. The different isoforms obtained by the protein modellling used are also shown regarding to the different aminoacid sequence. White boxes indicate the lack of the protein regarding to the complete structure. Boxes with inclined lines represents the amino acids insertions mainly due to the intron insertion in the transcribed mRNA.. See material and methods for software used.

Table 1. Amplification and sequencing primers used for general and specific amplification of LGP2 alternative transcripts.

713

Primer name	Sequence (5´→3´)
3UT-F1	CAACTGACTGCAGAAGGGC
3UT/E1C-F1	TCCATTTCTGATTTCTGCTCTCTG
E1-F3	TTCTGCTCTCTGCGCTGAG
E1-F4(0)	GATTTCTGCTCTCTGCGCTG
E1/I1.1-F1	CAAGGCGCAGGTTTTCAGT
E1/I1.24-F2	GGCGCAGGTGTAGCCTG
E1/I1.5-F1	AGGCGCAGGTGTAGCCTG
E1S-F1	GTGAGTCCTGGGCTGGG
E1/2-F2	AAGGCGCAGTTTCAGTTTCC
I1.2/E2-F2	GGCGATGACTTTTCAGTTTCC
I1.4-E2-F1	CCTTCCCTCAGTTTCAGTTTCC
I1.5/E3-F1	ATTCCCAGAATGGAGCTTCG
E2-F3:	GGGTCTCTGAGCTGAGCAGAG
E2-I2-F1	CTACCTACTAGAGCAGGTGAGTCCTG
E2/I2A-F1	CTACCTACTAGAGCAGGTGTAGCCTG
E2/3-F1:	CTACTAGAGCAGAATGGAGCTTCG
I2/E3-F1	ATTCCCAGAATGGAGCTTCG
I3/E3-F1	GTCAACAGGGTGAGTGGTCTG
E3-F1	GCAAGAATATCATCATCTGGCTG
E3/4-F2	GGTCAACAGGGTGCACCTG
E4-F1	ATCTGCACAGCAGAGCTTCTG
E4/5-F2	CGTGGAGCTCACTGTCTTCTC
E5-F1	ACCGTCTACAACGTCATCATGAG
E5/6-F1	GCAGCTCTGTGCCAACTTG
E6-F3	CCAACAGCCTTGCAAACAG
E6/7-F1	GCAGCCAGGATCCGTTTG
E7-F1	CAGGTGGTGAAGCTGAGTGAG
E7/8-F2	GAAGCTGAGTGAGGCTGCG
E8-F1	GAGCAACGGGTGTATGCG
E10-F1	AAGTTCCAAGATGGAACCCTG
E10/I10-F1	ATGGTCCAGGTACTCACATCCTC
I10/E11-F1	CTCTTCCCAGGCCAGGG
E10/E11-F1	TCCATGGTCCAGGCCAG

Primer name	Sequence (5´→3´)
E1S-R1	CCCAGCCCAGGACTCAC
E1/I1.5-R1	CAGGCTACACCTGCGCC
I1.5/E3-R1	CCATTCTGGGAATGGCAG
E2-R1	GCTCTAGTAGGTAGGTCTGCCCAG
E2/I2-R1	GGACTCACCTGCTCTAGTAGGTAGGTC
E2/I2A-R2	GGCTACACCTGCTCTAGTAGGTAGGTC
I2/E3-R1	TCCATTCTGGGAATGGCAG
E3-R1	ACCTTGGCTCCATCCACAG
E3/I3-R1	CAGACCACTCACCCTGTTGAC
E3/4-R1	CCAGGTGCACCCTGTTGAC
E4-R2	CATCTGCAGAAGCTCTGCTG
E5-R1	CATGATGACGTTGTAGACGGTG
E6-R1	CCTGTGGCAGAGGTTGTACTG
E7-R1	CTCACTCAGCTTCACCACCTG
E8-R1	ATAGAAATCCTGCAGCGCAG
E8/E9-R1	TCTTGCGGTCATCGAACAG
E8/E10-R1	CTGGTCCCTCATCGAACAG
E9/E10-R1	GGTCCCTCTGGGTCATGTG
E10-R1	GGGTTCCATCTTGGAACTTCTG
E10/I10-R1	TGTGAGTACCTGGACCATGGAG
I10/E11-R1	CCCTGGCCTGGGAAGAG
E10/E11-R1	TGGCCTGGACCATGGAG
E11-R1	GTCCATTTTCTGCACAGCAG
E11/E12-R1	CGGATCTTGGCCTGGTACTC
E13-R1	TGAAGACTTTGTTGATGACCACAG
E14-R1	AGTCAAAGTCAGGCACGGAG
E11/E13-R1	TATAGTAGTTCCTTGGCCTGGTACTC

714

715

716

717

718

719

720

721

722

723

Table 2. Percentage of homology of the intron 1 (A), intron 2 (B), and intron 3 (C) insertions in human and non-human primates LGP2 transcripts.

A)

INTRON 1 INSERTION	LARGE										SHORT				
	Patr-VX1 (n=392)	Patr-VX2 (n=392)	Patr-VX5 (n=392)	Patr-VX6 (n=392)	Papa-VX3 (n=396)	Hosa-V5 (n=399)	Hosa-V9 (n=399)	Hosa-VX3.1 (n=398)	Hosa-VX5 (n=398)	Hosa-VX3 (n=399)	Patr-VX4 (n=304)	Papa-VX1 (n=307)	Hosa-V4 (n=309)	Hosa-v8 (n=309)	Hosa-VX1.1 (n=308)
Patr-VX2 (n=392)	100														
Patr-VX5 (n=392)	100	100													
Patr-VX6 (n=392)	100	100	100												
Papa-VX3 (n=396)	99	99	99	99											
Hosa-V5 (n=399)	96	96	96	96	96										
Hosa-V9 (n=399)	96	96	96	96	96	100									
Hosa-VX3.1 (n=398)	96	96	96	96	96	99	99								
Hosa-VX5 (n=398)	96	96	96	96	96	99	99	99							
Hosa-VX3 (n=399)	96	96	96	96	96	99	99	99	99						
Patr-VX4 (n=304)	100	100	100	100	99	96	96	96	96	96					
Papa-VX1 (n=307)	99	99	99	99	100	96	96	96	96	96	99				
Hosa-V4 (n=309)	96	96	96	96	96	99	99	99	99	100	96	96			
Hosa-V8 (n=309)	96	96	96	96	96	99	99	99	99	100	96	96	100		
Hosa-VX1.1 (n=308)	96	96	96	96	96	99	99	99	99	99	96	96	99	99	
Hosa-VX1 (n=309)	96	96	96	96	96	99	99	99	99	100	96	96	100	100	99

B)

INTRON 2 INSERTION	LARGE			SHORT					
	Hosa-V6 (n=219)	Hosa-V4 (n=219)	Hosa-V5 (n=219)	Patr-VX5 (n=192)	Hosa-V2 (n=192)	Hosa-V3 (n=192)	Hosa-V7 (n=192)	Hosa-V8 (n=192)	Hosa-V9 (n=192)
Hosa-V4 (n=219)	100								
Hosa-V5 (n=219)	100	100							
Patr-VX5 (n=192)	98	98	98						
Hosa-V2 (n=192)	100	100	100	98					
Hosa-V3 (n=192)	100	100	100	98	100				
Hosa-V7 (n=192)	100	100	100	98	100	100			
Hosa-V8 (n=192)	100	100	100	98	100	100	100		
Hosa-V9 (n=192)	100	100	100	98	100	100	100	100	
Gogo-VX2 (n=192)	99	99	99	98	99	99	99	99	99

c) 736

INTRON 3 INSERTION	Patr-VX1 (n=227)	Patr-VX6 (n=227)
Patr-VX6 (n=227)	100	
Hosa-V3 (n=58)	81	81

737 738 739 740 741 742