Semin Immunopathol (2010) 32:437–444
DOI 10.1007/s00281-010-0224-x

REVIEW

Human *IRGM* gene "to be or not to be"

Cemaletin Bekpen · Ramnik J. Xavier · Evan E. Eichler

Received: 19 April 2010 / Accepted: 8 August 2010 / Published online: 25 August 2010

Abstract The immunity-related GTPases (IRG proteins) are one of the strongest early resistance systems against intracellular pathogens. The *IRG* gene family contains 21 copies arranged as tandem gene clusters on two chromosomes in the C57BL/6 mouse genome but has been reduced to only two copies in humans: *IRGC* and *IRGM*. *IRGC* is not involved in immunity, but the human *IRGM* gene plays a role in autophagy-targeted destruction of *Mycobacterium tuberculosis* (BCG) and *Salmonella typhimurium*. Variant *IRGM* haplotypes have been associated with increased risk for Crohn's disease and correlated with differential expression of *IRGM* transcripts. This article reviews in detail the studies performed on human samples, in vitro, and in sequence analyses that provide evidence for the unusual evolutionary history of the *IRGM* locus and the important role of the *IRGM* gene in autophagy and Crohn's disease in response to pathogenesis.

Keywords Human *IRGM* gene · *IRG* gene family · Autophagy · Crohn's disease · Innate immunity · Genome-wide association studies · Evolution of *IRGM* · Death and resurrection of a gene

This article is published as part of the Special Issue on Autophagy.

C. Bekpen · E. E. Eichler
Department of Genome Sciences, University of Washington,
Seattle, WA, USA

C. Bekpen · E. E. Eichler
Howard Hughes Medical Institute,
Seattle, WA, USA

R. J. Xavier
Center for Computational and Integrative Biology and
Gastrointestinal Unit, Massachusetts General Hospital,
Harvard Medical School,
Boston, MA, USA

R. J. Xavier
Broad Institute of MIT and Harvard,
Cambridge, MA, USA

Present Address:
C. Bekpen (✉)
Department of Molecular Biology and Genetics,
Boğaziçi University,
Bebek,
Istanbul 34342, Turkey
e-mail: bekpen@u.washington.edu
e-mail: cemalettin.bekpen@boun.edu.tr

Introduction

Immunity-related GTPases (IRG) are one of the strongest pathogen-resistance systems in mouse, but they can be found as multiple tandem copies in the genomes of most mammalian species. The gene family is induced by interferons in mouse and dog cells and has been implicated in innate resistance against a wide variety of intracellular pathogens including *Listeria monocytogenes*, *Toxoplasma gondii* [1, 2], *Mycobacterium tuberculosis* [3], *Salmonella typhimurium* [4], and *Chlamydia trachomatis* [4–9].

In the C57BL/6 mouse genome, *IRG* genes are organized as tandem gene clusters mapping to chromosome 11 and 18. *Irgc*, which is likely not associated with immune resistance, is found on chromosome 7. The *IRG* family consists of a total of 21 *IRG* genes in mouse, a few of which are probably or certainly pseudogenes [10, 11]. The IRG proteins have an N-terminal GTP-binding domain (G-domain) and a highly variable C-terminal region. The G-domain of the *IRG* family comprises all five classical GTP-binding motifs (Fig. 1) [12]. *IRG* genes have no homology to other GTPases except for the conserved G-domain. Both N- and C-terminal regions

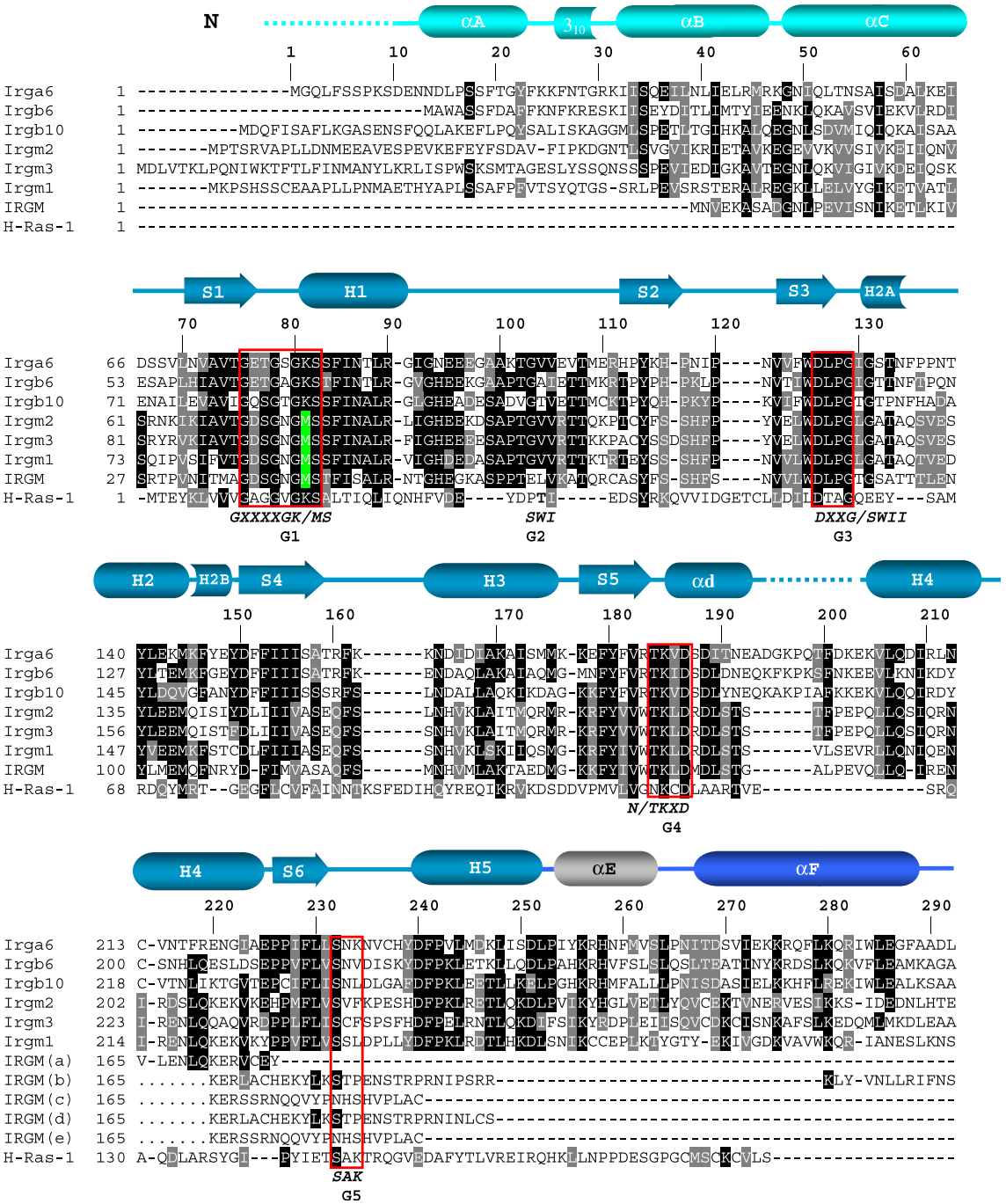

Fig. 1 Multiple sequence alignments of the GTPase domains of IRG proteins. Sequences of GTPase domains of IRG proteins, Irga6 (AJ007971), Irgb6 (L38444), Irgb10 (M63630), Irgm1 (U19119), Irgm2 (AJ007972), Irgm3 (U53219), human IRGM (ACF21844), and H-Ras-1 (P01112), showing close homology, aligned on the known secondary structures of Irga6 (IIGP1) [12]. Canonical GTPase motifs are indicated in *red boxes*. *IRGM* (*a*–*e*) splicing isoforms are presented after the splice region, and highlighted sequences in *green* in the G1 motif of GMS proteins indicate the unusual methionine residue that is unique for the IRG proteins. Multiple alignments are calculated using the server Clustal-W (EBI) and manually edited. The alignment is highlighted with *boxshade* server using the default options

have characteristic features that distinguish this family from other P-loop GTPases [13]. The IRG proteins can be grouped into two structural subfamilies, named GMS and GKS, based on an unusual amino acid substitution in the G1 motif (GX_4GK/MS) [14]. The GMS proteins, Irgm1, Irgm2, and Irgm3 (GMS subfamily), carry a methionine (M) instead of lysine (K) in their G1 motif (Fig. 1). This amino acid replacement (GKS to GMS) is a unique feature of the GMS type proteins [10]. All other P-loop GTPases have a canonical lysine (K) residue, which is important

for the coordination of the phosphates in bound GTP [13].

Interestingly, the family of *IRG* genes has been reduced to only two copies in the human genome and only the *IRGC* gene on chromosome 19 appears to be complete. *IRGC* is 89% identical at the protein level to mouse *Irgc*, the isolated member of the mouse GKS subfamily on chromosome 7, and is syntenic between the two species [10] (Rohde et al., manuscript in preparation). Neither murine *Irgc* nor human *IRGC* is induced by interferon, and mRNA expression is restricted to the testes in both species (Rohde et al., manuscript in preparation). The second human *IRG* gene, *IRGM*, is a fragment on human chromosome 5 in a region syntenic to both mouse chromosomes 18 and 11. *IRGM* encodes an N- and C-terminally truncated G-domain that is homologous to the mouse *Irgm* genes in the GMS subfamily and also carries the characteristic methionine in the G1 motif (see above and Fig. 1) [10].

The human *IRGM* gene

Human *IRGM* mRNA transcripts can be found in five different 3′-splicing isoforms (*IRGM a–e*), extending more than 30 kb 3′ of the long coding exon. By a combination of genomic and expressed sequence tag database analysis, in conjunction with 5′- and 3′-rapid amplification of cDNA ends (RACE) analysis from cultured human cell lines (HeLa and HEK 293), it was possible to clone different transcripts containing the G-domain of *IRGM* [10, 15] (Figs. 1 and 2). Yet none of the spliced transcripts extended beyond the G-domain of classical IRG proteins. Attempts to identify other splicing isoforms failed, perhaps due to the low level of *IRGM* mRNA expression in cultured cell lines. Therefore, it might be possible to find additional spliced forms of *IRGM*, especially if different human tissues are used. However, the protein products of the splice variants have predicted molecular weights of between 19 and 24 kD, and their expression at the protein level has not been documented. In most cell types studied to date, expression of *IRGM* transcripts is very low and detection of the endogenous protein has been difficult.

The shortest of the identified transcripts, *IRGM* (*a*), consists of two exons encoding 181 amino acids. *IRGM* (*a*) neglects a splice site immediately downstream of the open reading frame (ORF) and terminates at a polyadenylation signal sequence at the beginning of the second intron (Fig. 2). The longer transcripts, *IRGM* (*b*)–(*e*), include this splice site and splice the first exon to two or more downstream exons. *IRGM* transcripts have a highly unusual structure with an extended 5′-untranslated region (UTR), which contains Alu sequence (AluSc), endogenous retroviral nine (ERV9), and a 3′-UTR that includes alternatively spliced intronic sequence and exon–intron boundaries downstream of the putative termination codon in three of the four spliced forms *IRGM* (*c*)–(*e*). These latter would be expected to lead to rapid RNA degradation via nonsense-

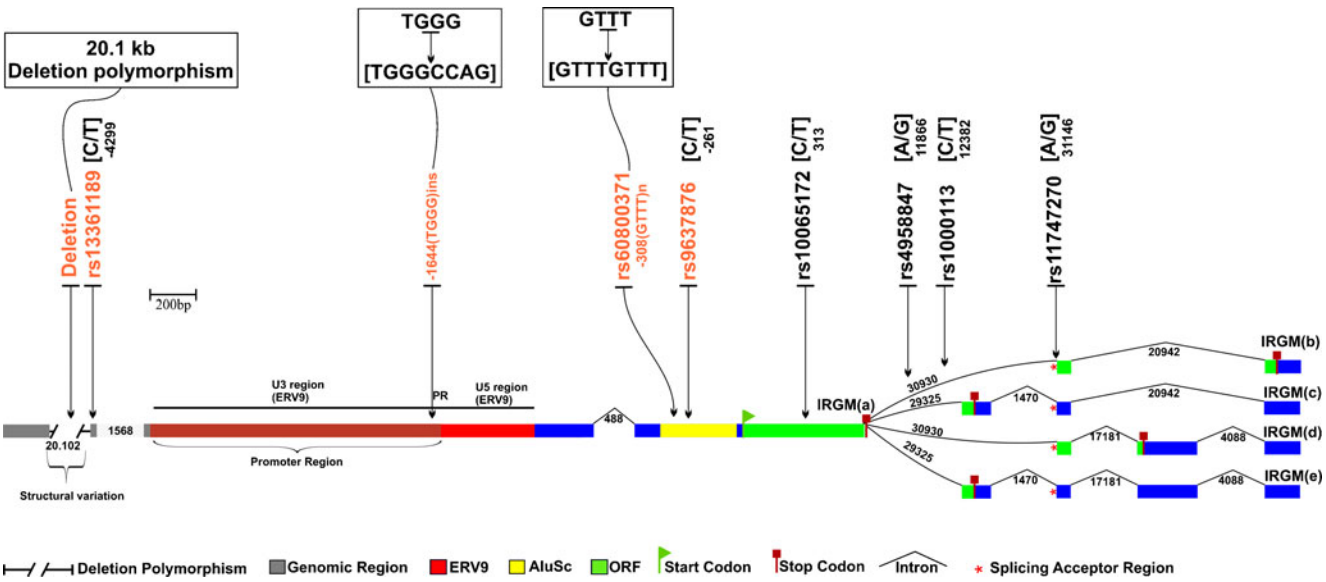

Fig. 2 A schematic summarizing the location of a sequenced structural variation and SNP polymorphisms with respect to the *IRGM* gene. Associations of the most significant SNPs at the *IRGM* locus are depicted. Deletion polymorphism (AC207974 from NA18956 [ABC9]) [15, 26], rs13361189 [19], −1644(TGGG)ins [29], rs60800371(−308(GTTT)n) [28, 29], rs9637876 [28], rs10065172 [26], rs4958847 [19], rs1000113 [29], and rs11747270 [68] are included. The positions of the *IRGM* variants [] are shown starting from the start codon (build 36/Hg18). The *orange*-colored SNPs indicate the variants that are shown to be associated with altered *IRGM* expression. The *figure* is scaled except for the genomic region (*gray*)

mediated decay (reviewed in detail in [16] and [17]; Fig. 2). However, polyadenylated mature mRNAs from all spliced isoforms were easily detectable by classical RT-PCR. This suggests that some other type of regulation might play a role in controlling the level of *IRGM* mRNA expression.

In contrast to mouse *IRG* genes, our experiments failed to show the induction of the human *IRGM* gene by interferons [10] (Bekpen, unpublished results). Therefore, although *IRGM* is very similar to the G-domains of the three mouse *Irgm* genes, it must be functionally different from all classical *IRG* genes by having truncations both in the N- and C-terminal regions and not being induced by interferons. This was one of the reasons why *IRGM* was initially thought to be pseudogene [10].

IRGM function in infection and association with pathogenicity

The first evidence indicating that *IRGM* might be a functional gene came from the Singh group [18], showing that *IRGM* is associated with the autophagy-targeted destruction of *Mycobacterium bovis*, BCG [18]. Genome-wide association scan studies showing a contribution of *IRGM* variants (rs13361189 and rs4958847) to Crohn's disease (CD) susceptibility followed this remarkable finding [19]. Furthermore, non-coding single nucleotide polymorphism (SNP) variants in the *IRGM* locus were shown to be associated with an increased risk for CD in different populations such as British [19, 20], German [21], New Zealander [22], Italian [23], Dutch, Belgian [24], and Spanish [25] (Fig. 2). Yet, the question of "how *IRGM* variants contribute to CD pathogenesis" remains unresolved. Interestingly, McCarroll and colleagues [26] showed that the *IRGM* SNP variant (rs13361189) was perfectly correlated (r^2=1.0) with a structural polymorphism 20.1 kb upstream of the human *IRGM* gene (Fig. 2). The structural variant deletion polymorphism was significantly associated with CD (172 cases and 344 controls, $p<$ 0.01) and the deletion allele was shown to be correlated with differential expression of *IRGM* in culture cells [15, 26]. They [26] also confirmed that reduction in the *IRGM* mRNA expression in culture cells was associated with an impairment of induction of autophagy and the clearance of intracellular pathogens (*S. typhimurium*) [27]. Most recent evidence indicates that the other variants of the *IRGM* locus and especially variations in the promoter region maybe correlated with differential expression (Fig. 2) [28, 29]. For example, in another study [28], *IRGM* variant rs9637876 (−261 T), which lies within the AluSc region, was significantly associated with increased levels of expression and contributes to protection from intracellular pathogen *M. tuberculosis* but not *Mycobacterium africanum* strains that are the cause of tuberculous disease. It is therefore reasonable to assume that the function of the *IRGM* gene might be directly or partially related to the level of *IRGM* mRNA expression. Furthermore, of most significant SNPs associated with CD, rs11747270 is located 280 bp upstream from the beginning of the fourth exon of *IRGM* (position 150.239.060 (build 36/hg18); Fig. 2). Given that this is very close to the splice-acceptor region and strongly associated with CD, it might be critical for spliceosome assembly and therefore determine the type of the spliced isoforms that are expressed. Additionally, one CD associated variant, which is a deletion allele upstream of *IRGM* [26], was shown to be correlated with altered expression of spliced isoforms of *IRGM* (Fig. 2) [15]. Thus, we can also suggest that alternative splicing of *IRGM* may be a mechanism critical for regulating the level of *IRGM* transcripts and thereby controlling pathogen loads. However, the role of *IRGM* variants associated with altered expression of *IRGM* in CD pathogenesis remains to be explained. Direct functional and molecular analyses are required to resolve the contribution of the *IRGM* variants to CD pathogenesis. Examples of such analyses have been performed for other immunity-related genes (e.g., NOD2 [30, 31], IL-10 [32], CYLD [33, 34], and ATG16L1 [35]), and the associated genetic variations were indeed shown to contribute to the development of intestinal diseases such as CD and ulcerative colitis. This contribution of *IRGM* genetic variants to CD or ulcerative colitis pathogenesis may be due to aberrant immune reactions or partial loss of immune tolerance to the intestinal commensal bacteria, but at present, there is no mechanistic insight into the basis of such effects.

Evolution of *IRGM*

Recently, we have shown that *IRGM* has had an unusual evolutionary history (Fig. 3) [15]. Briefly, our analyses suggest that the human *IRGM* gene was pseudogenized approximately 50 million years ago (mya) as a result of an Alu insertion event that disrupted its ORF. In all Old World and New World monkeys, the gene is non-functional. However, in the ancestor of apes and humans, function of *IRGM* was restored, presumably by the integration of an ERV9 element at a region 5′ upstream of the *IRGM* locus [15]. This is a rare case in primate evolution where a functional gene has been pseudogenized for millions of years and then brought back to function. In some species, such as the gibbon, evidence for both the functional and non-functional state exists [15, 36]. Nonetheless, it is still unclear when *IRGM* was truncated and lost its ability to be stimulated by interferon in the common ancestor of anthropoids in the primate lineage. Having a start codon at the very beginning of the long exon, a few bases

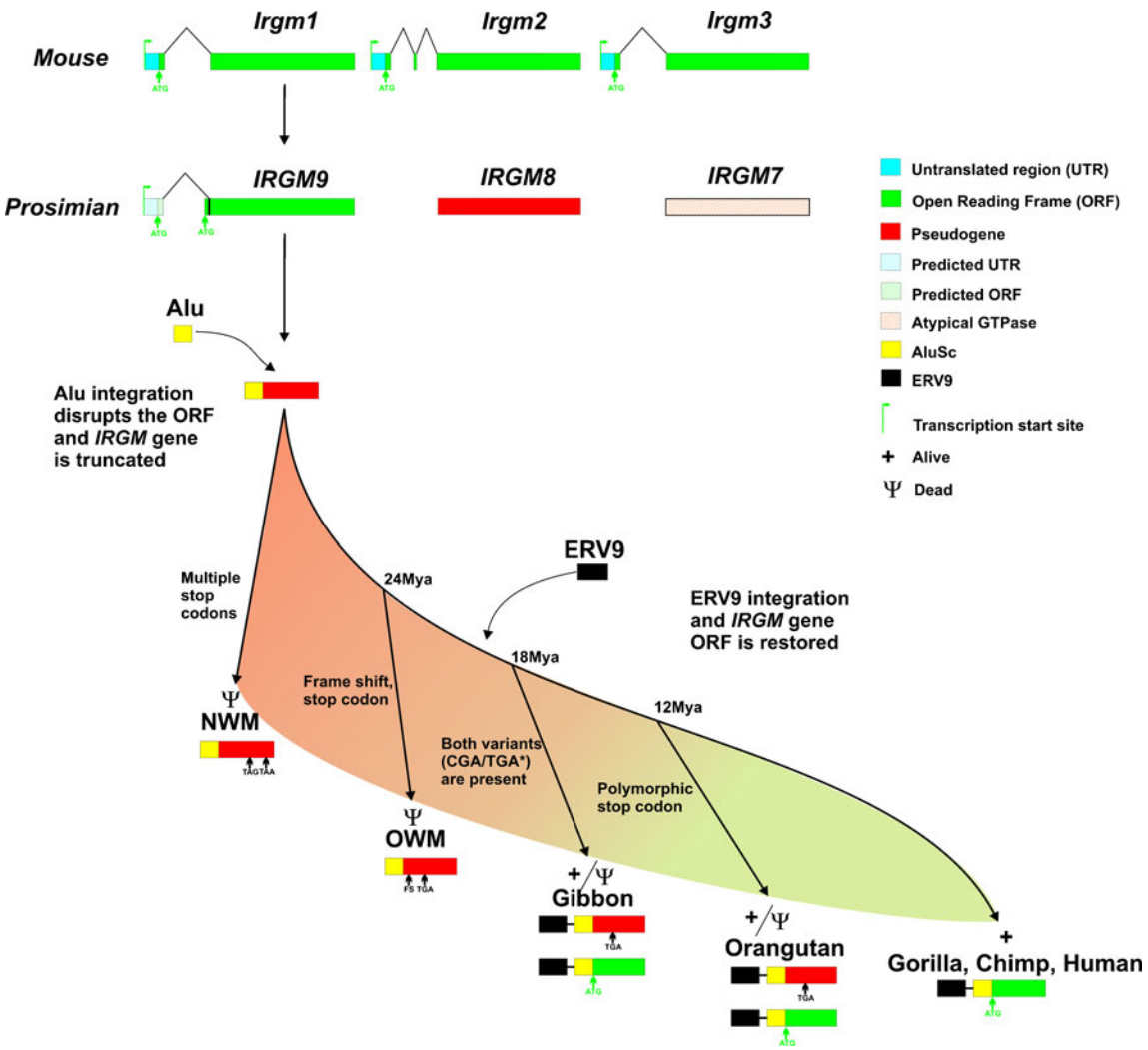

Fig. 3 The evolutionary history of the *IRGM* [15]. Approximately 50–60 mya the *IRG* family existed as a tandem gene family that is contracted to a single copy within the catarrhine lineage where the gene was pseudogenized in Old World (*OWM*) and New World monkey (*NWM*) species but then restored its open reading frame (*ORF*) in the human/ape lineage (with the exception of the orangutan species, which is polymorphic for both functional and pseudogenized versions of *IRGM*). Evolutionary and phylogenetic analyses support a model where the gene has been "dead" for at least 25 Ma of human primate evolution. This rebirth or restoration of the gene coincided with the insertion of an endogenous retrovirus that now serves as the functional promoter driving human gene expression. *Asterisk* of the five gibbon species analyzed, *H. gabriellae* shows a heterozygote stop codon. In human and African great ape, the functional copy becomes fixed. Frameshift mutation (Fs) and stop codons are indicated. The genomic loci are not drawn to scale with the exception of the full-length sequence of the *IRGM* ORF

downstream from the splice acceptor are typical of the *IRG* genes [10]. The ORF of *IRGM* is disrupted by the same AluSc sequence insertion in all anthropoids tested. Alu elements are the most abundant mobile elements in the human genome. Most Alu repeats in the primate lineage were duplicated more than 40 mya, with a rate of approximately one new Alu insertion in every primate birth during early evolution [37]. According to latest evidence from Price and colleagues [38], the main AluSc expansion in primate genome was observed between 32 and 40 mya, which is somewhat later event than the split of the anthropoids from prosimians. It is perhaps the AluSc insertion event at the early period of primate phylogeny that may be the reason why *IRGM* is truncated and has lost interferon stimulation.

As mentioned above, the beginning of the 5′-UTR of the *IRGM* transcript is similar to the U5 region of an ERV9 element. The promoter region corresponds to the ERV9 U3 long terminal repeats without interferon response elements [10, 15]. It is surprising to see that the ERV9 insertion event at 5′ region of the *IRGM* gene coincided with its resurrection [15]. ERV9 elements have been shown to play a role in regulating transcription, and ERV9-driven expression is very efficient in embryonic and hematopoietic cells

[39–41]. The function of ERVs in humans is not known. However, some data suggest that human ERVs may be involved in the prevention of infections with related exogenous retroviruses or act as pathological agents in certain autoimmune disorders [40, 42]. ERV9 is an endogenous retroviral element belonging to a family containing at least 14 different subfamilies and is specific to primates. The appearance of ERV9 was calculated to be as early as 40 mya. The main expansion in primates was observed at approximately 10–20 mya [43, 44]. Interestingly, the promoter and transcriptional structure of the *IRGM* gene is very similar to the *ZNF80* gene [45, 46]. Thus, it can be assumed that the ERV9 integration to the promoter region of human *IRGM* in the hominoid lineage happened during the expansion period of the retroviral element within the primate lineages.

Of note, there is a long-term (TGA/CGA) polymorphism in six out of 12 orangutan individuals that were studied and in one out of five gibbon species (*Hylobates gabriellae*). The polymorphism generates functional (CGA) and non-functional (TGA-stop) codons in *IRGM* gene copies in *Pongo pygmaeus* (Fig. 3). However, it should be noted that of the 12 orangutan individuals analyzed, none carried homozygous non-functional (TGA) alleles. This unusual situation might be explained by long-term balancing selection maintaining polymorphisms in the *IRGM* gene. For example, in *Arabidopsis thaliana* [47, 48], the plasma membrane protein, *RPM1*, is responsible for recognition of *Pseudomonas syringae* (pathogen for plants). Susceptible individuals do not have the coding region of *RPM1*, and both susceptibility (*RPM1*$^-$) and resistance alleles (*RPM1*$^+$) are present together worldwide within natural populations. Tian and colleagues generated independent transgenic lines expressing *RPM1* and showed that all the transgenic plants have fitness a loss of about 9% reduction in total seed production [49]. Similarly, *Mx1* is a resistance factor against a variety of viruses in mouse, and mice lacking the entire *Mx1* gene are susceptible to influenza viruses. Of all the standard laboratory mouse strains, only A2G and SL/NiA carry the *Mx1*$^-$ allele. However, wild mice possess the *Mx*$^+$ and *Mx*$^-$ alleles at roughly equal frequencies [50–52]. The present/absent polymorphism of *Mx1* suggests that expression of *Mx1*, like *RPM1*, might have high fitness cost. However, as with the *IRGM* gene, there is as yet no direct evidence for a fitness cost of the *Mx1* gene in mice.

Conclusion and perspective

The *IRG* family has periodically expanded to multiple members by segmental duplications, yet reduced to very few genes during the course of primate evolution [10]. It is very hard to prove whether the AluSc insertion or the fitness cost of IRG proteins was primarily responsible for the disappearance of the family. However, the three GMS proteins, Irgm1, Irgm2 and Irgm3, are shown to be essential regulators of the GTPase cycle of the GKS proteins [53]. Regulator effects of GMS proteins are so important that if one of the three GMS regulator proteins is absent, the GKS effector proteins form GTP-bound aggregates so that they can no longer perform their function of relocating to the *T. gondii* vacuole and initiating vacuolar disruption [54, 55]. It seems that all three GMS proteins must be present for normal functional behavior of the other 18 members of the *IRG* family (GKS proteins) [53]. Even more strikingly, loss of Irgm1 results in an interferon-dependent collapse of the lymphomyeloid system during infection, causing generalized immunoincompetence and an early death [56–58]. Thus, it is reasonable to suggest that the disruption of one of the GMS proteins (Irgm1, Irgm2, or Irgm3) by AluSc insertion was perhaps responsible for the extinction of the entire family of IRG proteins from anthropoids leading to human lineage at primate phylogeny.

Yet, the question of "How is the function of the *IRG* gene family replaced in humans?" remains to be explained. All the innate immune mechanisms, such as nitric oxide and oxygen radicals [59, 60], tryptophan depletion [61, 62], cation depletion [63], autophagy [64], and TLRs [65], are present in the mouse as well as in human. It is conceivable that one or more of the mechanisms listed above may fill the gap left by the loss of the *IRG* genes in man [66]. However, there are a variety of other gene families that are specifically expanded by segmental duplication in the primate lineage ([67] and Bekpen unpublished results). One may hypothesize that one of these primate or hominoid-specific gene families might replace the mechanism of *IRG* family in man. Future functional and evolutionary analyses are required to determine whether such an event arose during primate evolution.

Acknowledgment We are grateful to Jonathan C. Howard for his critical discussions and suggestions during the preparation of this manuscript, which helped us to refine our discussions. We thank Stefan Fuss for carefully editing the text, and we are also indebted to Hicham Bouabe and Can Alkan for their help and specific discussions.

References

1. Collazo CM, Yap GS, Sempowski GD, Lusby KC, Tessarollo L, Woude GFV, Sher A, Taylor GA (2001) Inactivation of LRG-47 and IRG-47 reveals a family of interferon {gamma}-inducible genes with essential, pathogen-specific roles in resistance to infection. J Exp Med 194:181–188
2. Taylor G, Collazo C, Yap G, Nguyen K, Gregorio T, Taylor L, Eagleson B, Secrest L, Southon E, Reid S, Tessarollo L, Bray M, McVicar D, Komschlies K, Young H, Biron C, Sher A, Vande

Woude G (2000) Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IGTP. PNAS 97:751–755
3. MacMicking J, Taylor GA, McKinney J (2003) Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science 302:654–659
4. Taylor GA (2004) p47 GTPases: regulators of immunity to intracellular pathogens. Nat Rev Immunol 4:100–109
5. Bernstein-Hanley I, Coers J, Balsara ZR, Taylor GA, Starnbach MN, Dietrich WF (2006) The p47 GTPases Igtp and Irgb10 map to the *Chlamydia trachomatis* susceptibility locus Ctrq-3 and mediate cellular resistance in mice. Proc Natl Acad Sci USA 103:14092–14097
6. Macmicking JD (2005) Immune control of phagosomal bacteria by p47 GTPases. Curr Opin Microbiol 8:74–82
7. Taylor GA (2007) IRG proteins: key mediators of interferon-regulated host resistance to intracellular pathogens. Cell Microbiol 9:1099–1107
8. Howard J (2008) The IRG proteins: a function in search of a mechanism. Immunobiology 213:367–375
9. Zhao YO, Rohde C, Lilue JT, Konen-Waisman S, Khaminets A, Hunn JP, Howard JC (2009) *Toxoplasma gondii* and the immunity-related GTPase (IRG) resistance system in mice: a review. Mem Inst Oswaldo Cruz 104:234–240
10. Bekpen C, Hunn JP, Rohde C, Parvanova I, Guethlein L, Dunn DM, Glowalla E, Leptin M, Howard JC (2005) The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. Genome Biol 6:R92
11. Shenoy AR, Kim BH, Choi HP, Matsuzawa T, Tiwari S, MacMicking JD (2007) Emerging themes in IFN-gamma-induced macrophage immunity by the p47 and p65 GTPase families. Immunobiology 212:771–784
12. Ghosh A, Uthaiah R, Howard J, Herrmann C, Wolf E (2004) Crystal structure of IIGP1: a paradigm for interferon-inducible p47 resistance GTPases. Mol Cell 15:727–739
13. Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 317:41–72
14. Boehm U, Guethlein L, Klamp T, Ozbek K, Schaub A, Fütterer A, Pfeffer K, Howard JC (1998) Two families of GTPases dominate the complex cellular response to interferon-γ. J Immunol 161:6715–6723
15. Bekpen C, Marques-Bonet T, Alkan C, Antonacci F, Leogrande MB, Ventura M, Kidd JM, Siswara P, Howard JC, Eichler EE (2009) Death and resurrection of the human IRGM gene. PLoS Genet 5:e1000403
16. Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51–74
17. Conti E, Izaurralde E (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol 17:316–325
18. Singh SB, Davis AS, Taylor GA, Deretic V (2006) Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science 313:1438–1441
19. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Cardon L, Mathew CG (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 39:830–832
20. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J (2008) Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 40(6):710–712
21. Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M, Schreiber S (2008) Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet 40:713–715
22. Roberts RL, Hollis-Moffatt JE, Gearry RB, Kennedy MA, Barclay ML, Merriman TR (2008) Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-based cohort. Genes Immun 9:561–565
23. Latiano A, Palmieri O, Cucchiara S, Castro M, D'Inca R, Guariso G, Dallapiccola B, Valvano MR, Latiano T, Andriulli A, Annese V (2009) Polymorphism of the IRGM gene might predispose to fistulizing behavior in Crohn's disease. Am J Gastroenterol 104:110–116
24. Weersma RK, Stokkers PC, Cleynen I, Wolfkamp SC, Henckaerts L, Schreiber S, Dijkstra G, Franke A, Nolte IM, Rutgeerts P, Wijmenga C, Vermeire S (2009) Confirmation of multiple Crohn's disease susceptibility loci in a large Dutch–Belgian cohort. Am J Gastroenterol 104:630–638
25. Palomino-Morales RJ, Oliver J, Gomez-Garcia M, Lopez-Nevot MA, Rodrigo L, Nieto A, Alizadeh BZ, Martin J (2009) Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: a meta-analysis approach. Genes Immun 10:356–364
26. McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ (2008) Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 40:1107–1112
27. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D et al (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464:713–720
28. Intemann CD, Thye T, Niemann S, Browne EN, Amanua Chinbuah M, Enimil A, Gyapong J, Osei I, Owusu-Dabo E, Helm S, Rusch-Gerdes S, Horstmann RD, Meyer CG (2009) Autophagy gene variant IRGM −261 T contributes to protection from tuberculosis caused by *Mycobacterium tuberculosis* but not by *M. africanum* strains. PLoS Pathog 5:e1000577
29. Prescott NJ, Dominy KM, Kubo M, Lewis CM, Fisher SA, Redon R, Huang N, Stranger BE, Blaszczyk K, Hudspith B, Parkes G, Hosono N, Yamazaki K, Onnie CM, Forbes A, Dermitzakis ET, Nakamura Y, Mansfield JC, Sanderson J, Hurles ME, Roberts RG, Mathew CG (2010) Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. Hum Mol Genet 19:1828–1839
30. Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF, Eckmann L, Karin M (2005) Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 307:734–738
31. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G, Flavell RA (2005) Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 307:731–734
32. Kuhn R, Lohler J, Rennick D, Rajewsky K, Muller W (1993) Interleukin-10-deficient mice develop chronic enterocolitis. Cell 75:263–274

33. Zhang G, Huang Y, Yan K, Li W, Fan X, Liang Y, Sun L, Li H, Zhang S, Gao M, Du W, Yang S, Liu J, Zhang X (2006) Diverse phenotype of Brooke–Spiegler syndrome associated with a nonsense mutation in the CYLD tumor suppressor gene. Exp Dermatol 15:966–970
34. Costello CM, Mah N, Hasler R, Rosenstiel P, Waetzig GH, Hahn A, Lu T, Gurbuz Y, Nikolaus S, Albrecht M, Hampe J, Lucius R, Kloppel G, Eickhoff H, Lehrach H, Lengauer T, Schreiber S (2005) Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. PLoS Med 2:e199
35. Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK, Kishi C, Kc W, Carrero JA, Hunt S, Stone CD, Brunt EM, Xavier RJ, Sleckman BP, Li E, Mizushima N, Stappenbeck TS, HWt V (2008) A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature 456:259–263
36. Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M (2010) Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. Genome Biol 11:R26
37. Deininger PL, Batzer MA (1999) Alu repeats and human disease. Mol Genet Metab 67:183–193
38. Price AL, Eskin E, Pevzner PA (2004) Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res 14:2245–2252
39. Svensson AC, Raudsepp T, Larsson C, Di Cristofano A, Chowdhary B, La Mantia G, Rask L, Andersson G (2001) Chromosomal distribution, localization and expression of the human endogenous retrovirus ERV9. Cytogenet Cell Genet 92:89–96
40. Lower R, Lower J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci USA 93:5177–5184
41. Bannert N, Kurth R (2006) The evolutionary dynamics of human endogenous retroviral families. Annu Rev Genomics Hum Genet 7:149–173
42. Patience C, Wilkinson DA, Weiss RA (1997) Our retroviral heritage. Trends Genet 13:116–120
43. Costas J, Naveira H (2000) Evolutionary history of the human endogenous retrovirus family ERV9. Mol Biol Evol 17:320–330
44. Ling J, Pi W, Bollag R, Zeng S, Keskintepe M, Saliman H, Krantz S, Whitney B, Tuan D (2002) The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. J Virol 76:2410–2423
45. Di Cristofano A, Strazullo M, Longo L, La Mantia G (1995) Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. Nucleic Acids Res 23:2823–2830
46. Di Cristofano A, Strazzullo M, Parisi T, La Mantia G (1995) Mobilization of an ERV9 human endogenous retroviral element during primate evolution. Virology 213:271–275
47. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. Nature 400:667–671
48. Tian D, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. Proc Natl Acad Sci USA 99:11525–11530
49. Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature 423:74–77
50. Staeheli P, Grob R, Meier E, Sutcliffe J, Haller O (1988) Influenza virus-susceptible mice carry *Mx* genes with a large deletion or a nonsense mutation. Mol Cell Biol 8:4518–4523
51. Haller O, Acklin M, Staeheli P (1987) Influenza virus resistance of wild mice: wild-type and mutant Mx alleles occur at comparable frequencies. J Interferon Res 7:647–656
52. Jin H, Yamashita T, Ochiai K, Haller O, Watanabe T (1998) Characterization and expression of the Mx1 gene in wild mouse species. Biochem Genet 36:311–322
53. Hunn JP, Koenen-Waisman S, Papic N, Schroeder N, Pawlowski N, Lange R, Kaiser F, Zerrahn J, Martens S, Howard JC (2008) Regulatory interactions between IRG resistance GTPases in the cellular response to *Toxoplasma gondii*. EMBO J 27:2495–2509
54. Martens S, Parvanova I, Zerrahn J, Griffiths G, Schell G, Reichmann G, Howard JC (2005) Disruption of *Toxoplasma gondii* parasitophorous vacuoles by the mouse p47-resistance GTPases. PLoS Pathog 1:e24
55. Papic N, Hunn JP, Pawlowski N, Zerrahn J, Howard JC (2008) Inactive and active states of the interferon-inducible resistance GTPase, Irga6, in vivo. J Biol Chem 283:32143–32151
56. Feng CG, Zheng L, Lenardo MJ, Sher A (2009) Interferon-inducible immunity-related GTPase Irgm1 regulates IFN gamma-dependent host defense, lymphocyte survival and autophagy. Autophagy 5:232–234
57. Feng CG, Weksberg DC, Taylor GA, Sher A, Goodell MA (2008) The p47 GTPase Lrg-47 (Irgm1) links host defense and hematopoietic stem cell proliferation. Cell Stem Cell 2:83–89
58. Hunn JP, Howard J (2010) The mouse resistance protein, Irgm1 (LRG-47): a regulator or an effector of pathogen defense? PLoS Pathog 6(7):e1001008
59. Nathan C, Shiloh MU (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proc Natl Acad Sci USA 97:8841–8848
60. Fang FC (2004) Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. Nat Rev Microbiol 2:820–832
61. Pfefferkorn ER (1984) Interferon gamma blocks the growth of *Toxoplasma gondii* in human fibroblasts by inducing the host cells to degrade tryptophan. Proc Natl Acad Sci USA 81:908–912
62. Robinson CM, Shirey KA, Carlin JM (2003) Synergistic transcriptional activation of indoleamine dioxygenase by IFN-gamma and tumor necrosis factor-alpha. J Interferon Cytokine Res 23:413–421
63. Schaible UE, Kaufmann SH (2004) Iron and microbial infection. Nat Rev Microbiol 2:946–953
64. Gutierrez MG, Master SS, Singh SB, Taylor GA, Colombo MI, Deretic V (2004) Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. Cell 119:753–766
65. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A (2005) The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci USA 102:9577–9582
66. Nelson DE, Virok DP, Wood H, Roshick C, Johnson RM, Whitmire WM, Crane DD, Steele-Mortimer O, Kari L, McClarty G, Caldwell HD (2005) Chlamydial IFN-gamma immune evasion is linked to host infection tropism. Proc Natl Acad Sci USA 102:10658–10663
67. Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE (2007) Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nat Genet 39:1361–1368
68. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40:955–962