



OPEN

## Germline variants of *ATG7* in familial cholangiocarcinoma alter autophagy and p62

Stephanie U. Greer<sup>1,10</sup>, Jiamin Chen<sup>1,10</sup>, Margret H. Ogmundsdottir<sup>2,10</sup>, Carlos Ayala<sup>3</sup>, Billy T. Lau<sup>1</sup>, Richard Glenn C. Delacruz<sup>4,5</sup>, Imelda T. Sandoval<sup>4,5</sup>, Sigrun Kristjansdottir<sup>6</sup>, David A. Jones<sup>4,5</sup>, Derrick S. Haslem<sup>4</sup>, Robin Romero<sup>4</sup>, Gail Fulde<sup>4</sup>, John M. Bell<sup>7</sup>, Jon G. Jonasson<sup>6,8</sup>, Eirikur Steingrimsson<sup>9</sup>, Hanlee P. Ji<sup>1,7</sup>✉ & Lincoln D. Nadauld<sup>4</sup>✉

Autophagy is a housekeeping mechanism tasked with eliminating misfolded proteins and damaged organelles to maintain cellular homeostasis. Autophagy deficiency results in increased oxidative stress, DNA damage and chronic cellular injury. Among the core genes in the autophagy machinery, *ATG7* is required for autophagy initiation and autophagosome formation. Based on the analysis of an extended pedigree of familial cholangiocarcinoma, we determined that all affected family members had a novel germline mutation (c.2000C>T p.Arg659\* (p.R659\*)) in *ATG7*. Somatic deletions of *ATG7* were identified in the tumors of affected individuals. We applied linked-read sequencing to one tumor sample and demonstrated that the *ATG7* somatic deletion and germline mutation were located on distinct alleles, resulting in two hits to *ATG7*. From a parallel population genetic study, we identified a germline polymorphism of *ATG7* (c.1591C>G p.Asp522Glu (p.D522E)) associated with increased risk of cholangiocarcinoma. To characterize the impact of these germline *ATG7* variants on autophagy activity, we developed an *ATG7*-null cell line derived from the human bile duct. The mutant p.R659\* *ATG7* protein lacked the ability to lipidate its LC3 substrate, leading to complete loss of autophagy and increased p62 levels. Our findings indicate that germline *ATG7* variants have the potential to impact autophagy function with implications for cholangiocarcinoma development.

Autophagy is a conserved cellular pathway tasked with the recycling of misfolded proteins and damaged organelles. Playing a fundamental role in cellular homeostasis, autophagy regulates many different biological functions including the initiation of senescence and maintenance of genome stability. Reflecting its important intracellular function across many organisms, the core set of autophagy related genes (*ATG*) are highly conserved across mammals. Given their essential role in mammalian cell maintenance, autophagy genes are increasingly being implicated in the pathogenesis of different diseases that include neurodegenerative disorders, autoimmune diseases and cancer<sup>1</sup>.

Cellular stressors can initiate autophagy through various signaling pathways. The core *ATG* proteins mediate the induction and formation of autophagosomes allowing subsequent fusion with lysosomes and the degradation of autophagosomal cargo<sup>2</sup>. One of the critical steps following autophagy induction is the formation of a double membrane structure, known as the phagophore. Subsequently, lipidated LC3/GABARAPs are added to the phagophore and growing autophagosome where they recruit interacting partners. An example is the autophagy receptor protein p62/SQSTM1, also known as Sequestosome-1. This protein is recruited via a physical interaction to the autophagosome facilitating the degradation of ubiquitinated cargo during autophagy. Importantly, genetic

<sup>1</sup>Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Department of Anatomy, Faculty of Medicine, BioMedical Center, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. <sup>3</sup>Division of General Surgery, Department of Surgery, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>4</sup>Intermountain Precision Genomics Program, Intermountain Healthcare, Saint George, UT 84790, USA. <sup>5</sup>Oklahoma Medical Research Foundation, Oklahoma University, Oklahoma City, OK 73104, USA. <sup>6</sup>Department of Pathology, Landspítali-University Hospital, 101 Reykjavik, Iceland. <sup>7</sup>Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA. <sup>8</sup>Faculty of Medicine, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. <sup>9</sup>Department of Biochemistry and Molecular Biology, Faculty of Medicine, BioMedical Center, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. <sup>10</sup>These authors contributed equally: Stephanie U. Greer, Jiamin Chen and Margret H. Ogmundsdottir. ✉email: genomics\_ji@stanford.edu; Lincoln.Nadauld@imail.org

variants in ATG genes have functional implications in autophagy and lead to specific diseases<sup>1</sup>. For example, germline mutations among several autophagy genes, including *ATG5* and *ATG16L1*, have been identified among hereditary disorders involving neurodegeneration and autoimmunity. Although the role of autophagy in cancer remains unclear, SNPs in *ATG5* and *ATG7* have been associated with clear cell renal cell carcinoma<sup>3</sup>. Presumably, these mutations affect autophagy function.

Cholangiocarcinoma (CCA; MIM: 615619) is an aggressive, difficult-to-treat epithelial carcinoma arising from the bile duct. Once diagnosed, this biliary duct cancer has a dismal five-year survival rate of less than 10%<sup>4,5</sup>. CCAs are categorized into intrahepatic, perihilar and distal subtypes based on their anatomical location. Perihilar CCA, or Klatskin tumor, is anatomically located at the confluence of the right and left hepatic bile ducts, and accounts for approximately 50% of all cases<sup>6</sup>. Genomic studies of CCA have implicated a variety of known cancer drivers such as *TP53*, *KRAS*, *ERBB2* and others that play a role in oncogenic processes<sup>7–10</sup>.

In this study, we identified a novel germline mutation in *ATG7* (NM\_001349232.2: c.2000C>T p.Arg659\* (p.R659\*)) from a pedigree with familial CCA by comprehensive whole genome and exome sequencing analysis. This germline mutation segregated with affected family members all of whom had perihilar cholangiocarcinoma. Moreover, we detected somatic deletion of *ATG7* and loss of heterozygosity among the cancers in this family. Together, these mutations point to a complete loss of the wildtype alleles of *ATG7*. In addition, we conducted a population study of *ATG7* polymorphisms associated with CCA in the Icelandic population and identified a polymorphic variant of *ATG7* (NM\_001349232.2: c.1591C>G p.Asp522Glu (p.D522E)) that was associated with elevated risk of developing this cancer. For this study, we sought to determine the consequences of this mutation and polymorphism.

During autophagy, cytoplasmic materials targeted for degradation are encapsulated in a transient structure called an autophagosome, which subsequently fuses with a lysosome containing digestive enzymes<sup>11</sup>. Lipidated LC3 (LC3-II) is essential for autophagosome maturation and also for the recognition of targeted cargo in selective autophagy<sup>12,13</sup>. The *ATG7* protein plays an essential role during autophagosome formation because it is required for two critical conjugation reactions that are part of the multistep process to lipidate LC3. With its E1-like enzymatic activity, *ATG7* adenylates LC3-I so that it can be lipidated to generate LC3-II, and also *ATG7* adenylates *ATG12* as part of the process that generates *ATG5-ATG12* conjugates<sup>14,15</sup>. Genetically engineered mouse models initially demonstrated that hepatocyte-specific knockout of *ATG7* resulted in the formation of liver adenomas that did not progress to cancerous tumors<sup>16</sup>. However, a later study found that hepatocyte-specific knockout of *ATG7* resulted in hepatomegaly and hepatocellular carcinoma, suggesting autophagy deficiency might promote tumorigenesis in hepatobiliary tissues<sup>17</sup>.

To study the functional impact of the mutation p.R659\* and p.D522E polymorphism on autophagy function we conducted experiments using a human bile duct cell line. Based on the context, we hypothesized that the germline mutation would have greater functional consequences than the polymorphism. Our functional studies demonstrated that the *ATG7* germline mutation (p.R659\*) led to a loss of function in autophagy and increased p62 levels in bile duct epithelial cells. The *ATG7* polymorphism (p.D522E) had a less dramatic impact on autophagy function. Overall, these results suggest a potential role of autophagy in CCA development.

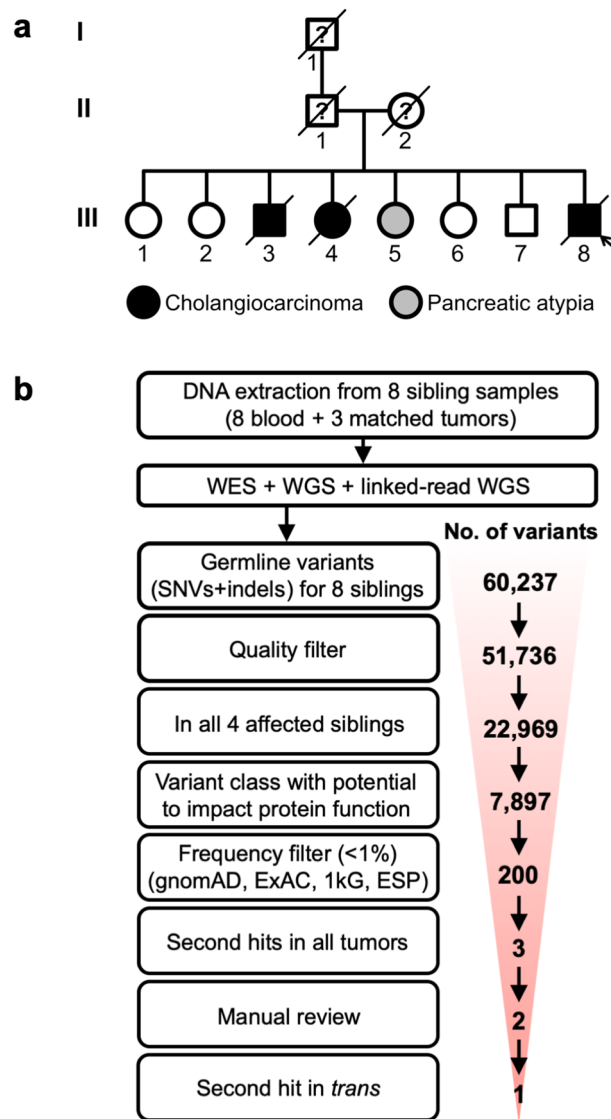
## Results

**Identifying a germline *ATG7* mutation associated with hereditary CCA.** We identified a family with a high incidence of perihilar CCA (Supplementary Table S1 and Fig. 1a). As noted previously, perihilar CCA was present in three of eight siblings (III:3, III:4, III:8) in generation III of the family. The pathology review from the tumors taken at surgical resection and CT imaging studies confirmed this diagnosis among these siblings. One of the eight siblings was diagnosed with pancreatic atypia (III:5). The average age at diagnosis in generation III was approximately 60 years. In addition, gastrointestinal cancers affected all three generations of this family. Specifically, the parents and the paternal grandfather were reported to have cancers related to the GI tract. The father (II:1) was reported to have had pancreatic and prostate cancer at the time of death, and the mother (II:2) had pancreatic and cecal tumors. The family also reported that the paternal grandfather (I:1) was diagnosed with stomach cancer. We enrolled and collected blood from all siblings in generation III of the family, both affected and unaffected.

Supplementary Fig. S1 outlines our approach for identifying candidate genes and germline mutations that may account for the increased incidence of CCA in this family. We conducted both exome and whole genome analysis including studies with linked reads that provided haplotype information. Sequencing and alignment metrics are available in Supplementary Table S2 and phasing metrics are available in Supplementary Table S3. First, we examined the germline exome variants of the eight siblings. Overall, we identified on average 33,203 single nucleotide variants (SNVs) per sibling and 3,994 insertion-deletions (indels) per sibling (Supplementary Table S4). When we combined the variants of all individuals, there were a total of 46,133 and 5,603 high quality SNVs and indels, respectively, among the eight siblings.

To identify germline mutations associated with the affected family members, we used the following criteria: the germline variant was present in all of the affected siblings albeit it could be present in unaffected siblings as well, assuming incomplete penetrance; the variant had the potential to alter protein function (i.e. the variant class had to be one of: missense, nonsense, splice site, stop lost, frameshift, in-frame); the variant had to occur at low frequency (i.e. less than 1%) in the 'healthy' population. This germline filtering strategy is similar to those used in other WES studies to identify putative causal variants in Mendelian disorders<sup>18</sup>.

Approximately 44% (22,969) of the variants we identified were present in all four of the affected siblings (Supplementary Table S5 and Fig. 1b). Of the variants in the affected siblings, 34% (N = 7897) had the potential to impact protein function. For the SNVs, there were 6,610 nonsynonymous variants, 656 splice site variants, and 75 variants affecting stop codons (Supplementary Table S6). For the indels, there were 203 in-frame variants,



**Figure 1.** Identification of a putative germline predisposition variant in a family with inherited CCA. **(a)** This pedigree depicts a family with an inherited predisposition to biliary cancers. The current study had access to samples from generation III of the family. **(b)** The number of putative germline variants (single nucleotide variants and insertion-deletions) is indicated for each stage of filtering.

191 frameshift variants, and 162 splice site variants (Supplementary Table S6). We cross-validated the variants with the ESP, ExAC and gnomAD databases, which allowed us to eliminate population polymorphisms that were unlikely to be the cause of a rare hereditary condition<sup>19</sup>. In total, 43 SNVs and 157 indels remained as putative germline mutations, present in 148 genes (Supplementary Table S5 and Fig. 1b).

The germline analysis of this family did not identify mutations in any of the previously reported CCA and pancreaticobiliary cancer predisposition genes. Familial cases of CCA have been reported in Lynch syndrome (MIM: 120,435) caused by mutations in the DNA mismatch repair genes which include *MLH1*, *MSH2*, *MSH6*, *PMS1*, and *PMS2*<sup>20</sup> and in *BAP1* tumor predisposition syndrome (MIM: 614,327), where the *BAP1* tumor suppressor gene is mutated<sup>21,22</sup>. Genetic studies of hereditary pancreatic cancers have included CCA given the similarities of their embryonic tissue origins and epithelial features which involve secretory physiologic functions<sup>7,23</sup>. These genes include *BRCA1*<sup>24</sup>, *BRCA2*<sup>25</sup>, *PALB2*<sup>26</sup>, *ATM*<sup>27</sup>, *STK11*<sup>28</sup>, *CDKN2A*<sup>29</sup>, and *BRIP1*<sup>30</sup>. We did not find mutations in any of the above-mentioned genes or in genes with mutations reported as risk factors for pancreatic cancer from recent genome-wide association studies (*TNS3*, *NOC2L*, *HNF4G*, *HNF1B*, *GRP*, *UHMK1*, *APIG2*, *DNTA*, *CHST6*, *FGFR3* and *EPHA1*)<sup>27,31–33</sup>. In addition, we used whole genome sequencing to identify germline deletions, rearrangements or other SVs that may be segregating with the affected individuals. We determined that there were no germline rearrangements associated with the affected individuals that implicated a susceptibility locus (Supplementary Table S7).

**Characterization of somatic mutations in CCA.** We evaluated whether any of the 148 candidate genes identified in our germline analysis also harbored somatic alterations in the tumors of the three siblings with CCA, thus indicating putative biallelic genetic events. We conducted whole exome and whole genome sequencing on three tumor-normal sample pairs (patient III:3, III:4, III:8) to identify somatic cancer mutations and other genetic aberrations. We examined somatic SNVs and indels to identify putative inactivating mutations that could represent second hits (Tables S8–S11). We also examined genomic copy number changes as a way of ascertaining whether any deletions overlapped with our candidate germline mutations. The tumor of patient III:3 had five somatic copy number deletions with an average size of 49.8 Mb and containing a total of 3050 genes. In patient III:4, there were 32 somatic deletions with an average size of 34.8 Mb containing 9,620 genes. Patient III:8 had 20 somatically deleted regions with an average size of 24.9 Mb and containing 3,053 genes (Supplementary Table S12).

We identified three candidate germline variants in genes that also harbored somatic mutations among the affected individuals after filtering (Supplementary Table S13). The first was a missense mutation in *SPSB1* (NM\_025106.4: c.945C>T p.Arg202Trp (p.R202W)) present in all eight siblings. The second was a nonsense mutation in *ATG7* (c.2000C>T [p.R659\*]) present in four affected siblings as well as two unaffected siblings (III:2 and III:6). The third was a missense mutation in the *PP2D1* gene (NM\_001252657.2: c.1898G>A p.Pro547Leu (p.P547L)) but it was eliminated as a candidate due to a lack of functional evidence based on variant impact predictions. The *PP2D1* substitution was scored as being benign and tolerant as per PolyPhen and SIFT analysis, respectively. Also, this mutation had a low CADD score (5.309) denoting low predicted pathogenicity (Supplementary Table S14). For *SPSB1* and *ATG7*, we identified both germline mutations and somatic alterations among all three affected individuals' tumors (Tables S13 and S14). Specifically, all three patient tumor samples harbored somatic copy number deletions that included these two genes (Tables S12 and S13).

**Linked read sequencing confirmed second hit for *ATG7*.** Next, we determined whether genomic deletions of the two candidate genes led to the loss of the remaining wildtype (WT) allele in the tumor sample. For this haplotype analysis, we applied linked-read sequencing to the tumor of patient III:8 to ascertain whether the somatic copy number deletion affected the mutant allele or the wildtype allele. We applied a chromosomal haplotyping method to link phase blocks<sup>34</sup> and identified 14 chromosome arms where extended haplotypes covered anywhere from 1 to 98% of a given chromosome arm (Supplementary Table S15). *SPSB1* is located on chromosome arm 1p in an encompassing 36.6 Mb haplotype ( $p = 5.3 \times 10^{-13}$ ). *ATG7* is located on chromosome arm 3p in an encompassing 44.4 Mb haplotype ( $p = 2.2 \times 10^{-16}$ ) (Supplementary Table S15). To visualize the haplotype structure across these genomic deletions, we plotted the two haplotypes such that individual phase blocks are color-coded according to their assignment to an extended genomic haplotype (Figs. 2a and S2).

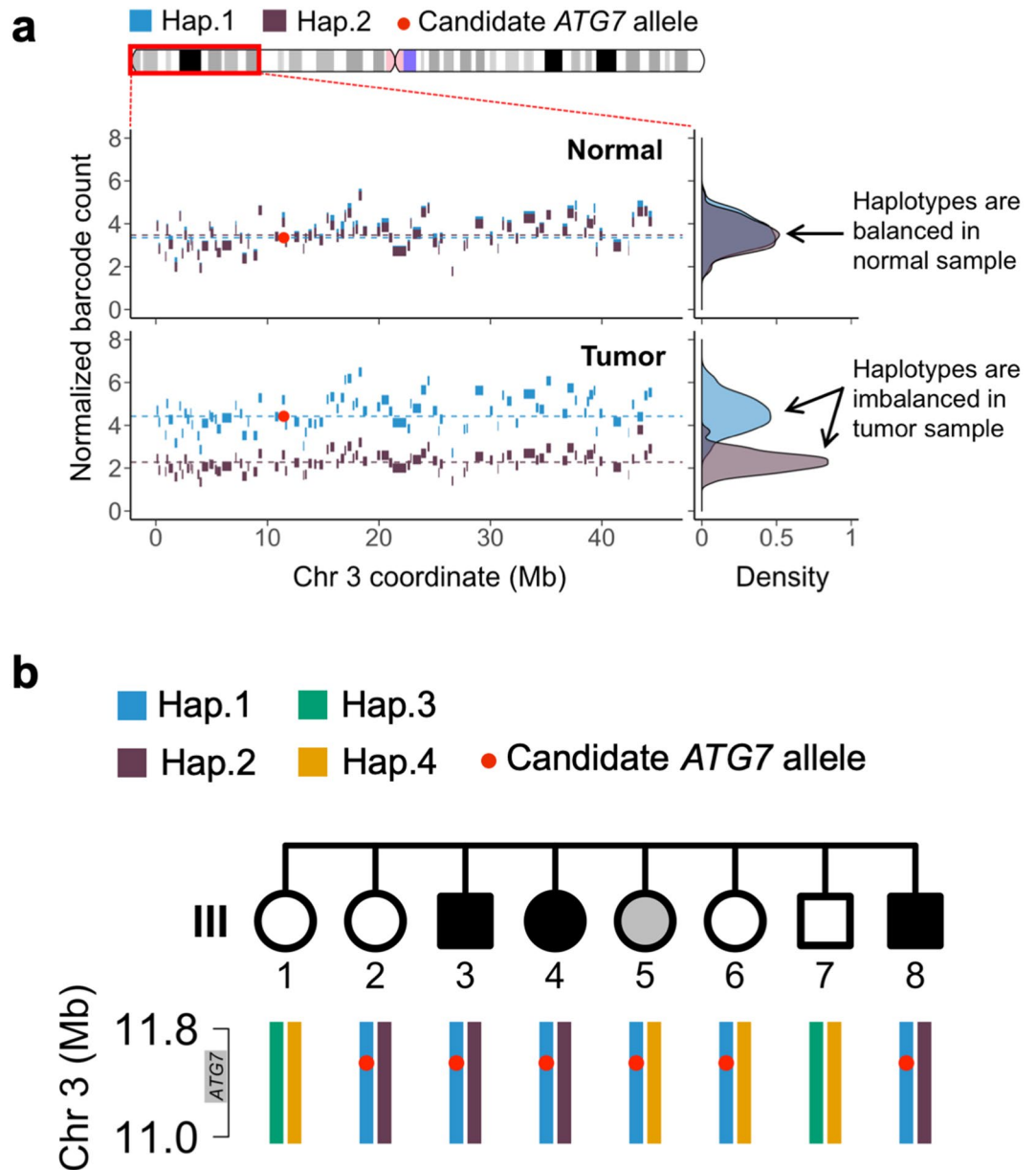
In the case of *SPSB1*, we discovered that the germline mutation (p.R202W) was deleted in the tumor genome and thus the remaining allele was wildtype (see Supplementary Table S16 and Supplementary Fig. S2 online). Based on the retention of the wildtype allele, we eliminated it as a possible tumor suppressor. Conversely, in the case of *ATG7*, we found that the wildtype *ATG7* allele was deleted in the tumor and thus the haplotype with the *ATG7* germline mutation (p.R659\*) was retained, leading to loss of heterozygosity (LOH) in tumor (Supplementary Table S16 and Fig. 2a). Together, the germline mutation and genomic deletion identified an *ATG7* loss of function among the CCAs from the affected individuals. Using the extended haplotype information from the germline analysis of all generation III siblings, we confirmed that all affected individuals had inherited the extended haplotype with the candidate p.R659\* allele in the *ATG7* genomic region (Fig. 2b).

**Identification of an *ATG7* polymorphism associated with CCA.** We leveraged the Icelandic population for a study of *ATG7* association with CCA. We analyzed sequence variants in *ATG7* identified in whole genome sequence data from 8,453 Icelanders sequenced to a median depth of 32 ×<sup>35</sup>. Based on the SNVs and single nucleotide polymorphisms (SNPs) that were identified from the sequencing data, we imputed the *ATG7* genotypes of an additional 150,656 Icelanders with SNP array information. Additional familial imputation, using the nationwide Icelandic genealogical database, allowed these genotypes to be propagated into 294,212 un-genotyped close relatives of the chip-genotyped individuals. Thus, the overall sample size was adequately powered to detect the association of specific *ATG7* variants and polymorphisms with an increased risk of developing CCA.

Six coding variants in *ATG7* reached our threshold of imputation quality (Supplementary Table S17). Using information from the Icelandic Cancer Registry (ICR), we tested the association between these variants and affected individuals with CCA ( $N = 353$ ). This analysis revealed an association between the minor allele of rs146589465 (rs146589465-G) and increased risk of CCA (OR 6.56,  $P = 1.3 \times 10^{-3}$ ; Table 1). We assessed the quality of the imputation of rs146589465 by directly genotyping imputed carriers ( $N = 154$ ) and non-carriers ( $N = 1180$ ). The concordance between imputed and directly measured genotypes was 0.999 (Supplementary Table S18).

The variant rs146589465-G results in an aspartic acid to glutamic acid substitution at position 522 in the *ATG7* protein (p.D522E). The variant is rare and is present in 1/400 individuals in the Icelandic population (allelic frequency is 0.12%). It is also present at a similar frequency in individuals of European ancestry per the gnomAD database (Supplementary Table S17). It is present only at a very low frequency in other populations, with its highest frequency being ~0.1% in the Latino population. To assess whether rs146589465 was associated with increased risk for other cancer types, we tested 19 additional tumor types using information from the ICR (Supplementary Table S19). The risk of hepatocellular carcinoma (OR = 4.95,  $P = 2.9 \times 10^{-2}$ ; Table 1) and pituitary adenoma ( $N = 422$ ; OR = 5.22,  $P = 3.9 \times 10^{-3}$ ) was also elevated with rs146589465-G (Table 1).

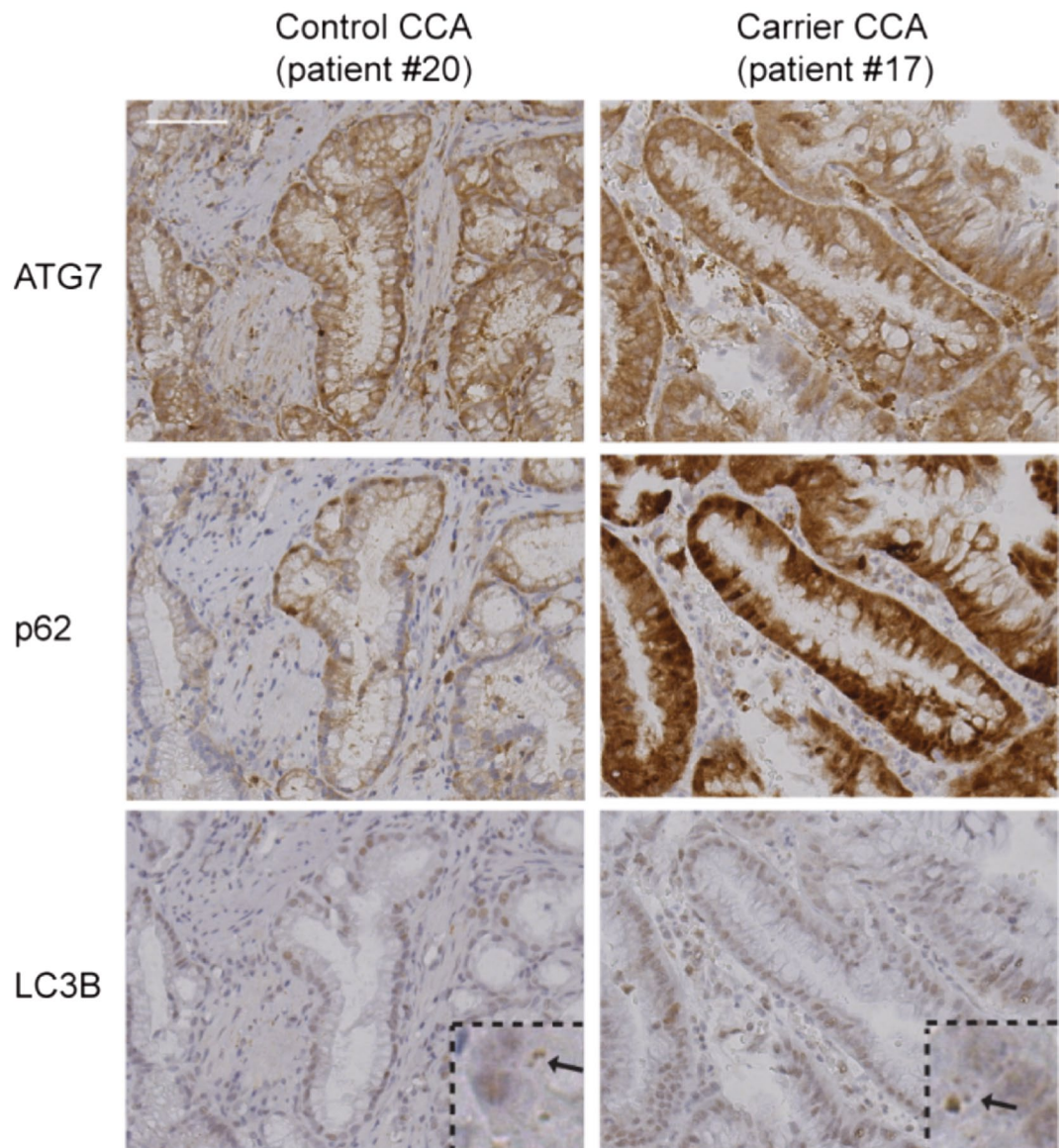
To evaluate the association of the *ATG7* rs146589465-G (p.D522E) variant with autophagy defects in patient samples, we analyzed CCA as well as hepatocellular tumors from 6 carriers and 22 non-carriers (Tables S20 and



**Figure 2.** Haplotype analysis of the candidate *ATG7* allele (p.R659\*). (a) Extended haplotype of the 45 Mb deleted region of chromosome 3p in individual III:8. The blocks indicate the original fragmented haplotypes, and their color denotes their subsequent assignment to haplotypes covering many Megabases. The candidate *ATG7* allele exists in haplotype 1 (blue), which was the non-deleted haplotype in the tumor of this individual. (b) Germline haplotype analysis of siblings for the *ATG7* genomic region. The haplotype segregation across all eight siblings in generation III was determined for the 0.8 Mb genomic region surrounding the *ATG7* allele (red dot). The candidate *ATG7* allele exists in haplotype 1 (blue).

A1 <sup>a</sup>	A2 <sup>a</sup>	Freq. A1 <sup>a</sup> [%]	Phenotype	N cases	N controls	P-value	OR <sup>a</sup> (A1)	95% CI
G	C	0.12	Cholangiocarcinoma	353	233,169	1.3 × 10 <sup>-3</sup>	6.56	(2.085–20.64)
			Pituitary adenoma	422	303,642	3.9 × 10 <sup>-3</sup>	5.22	(1.699–16.03)
			Hepatocellular carcinoma	280	301,578	2.9 × 10 <sup>-2</sup>	4.95	(1.178–20.80)

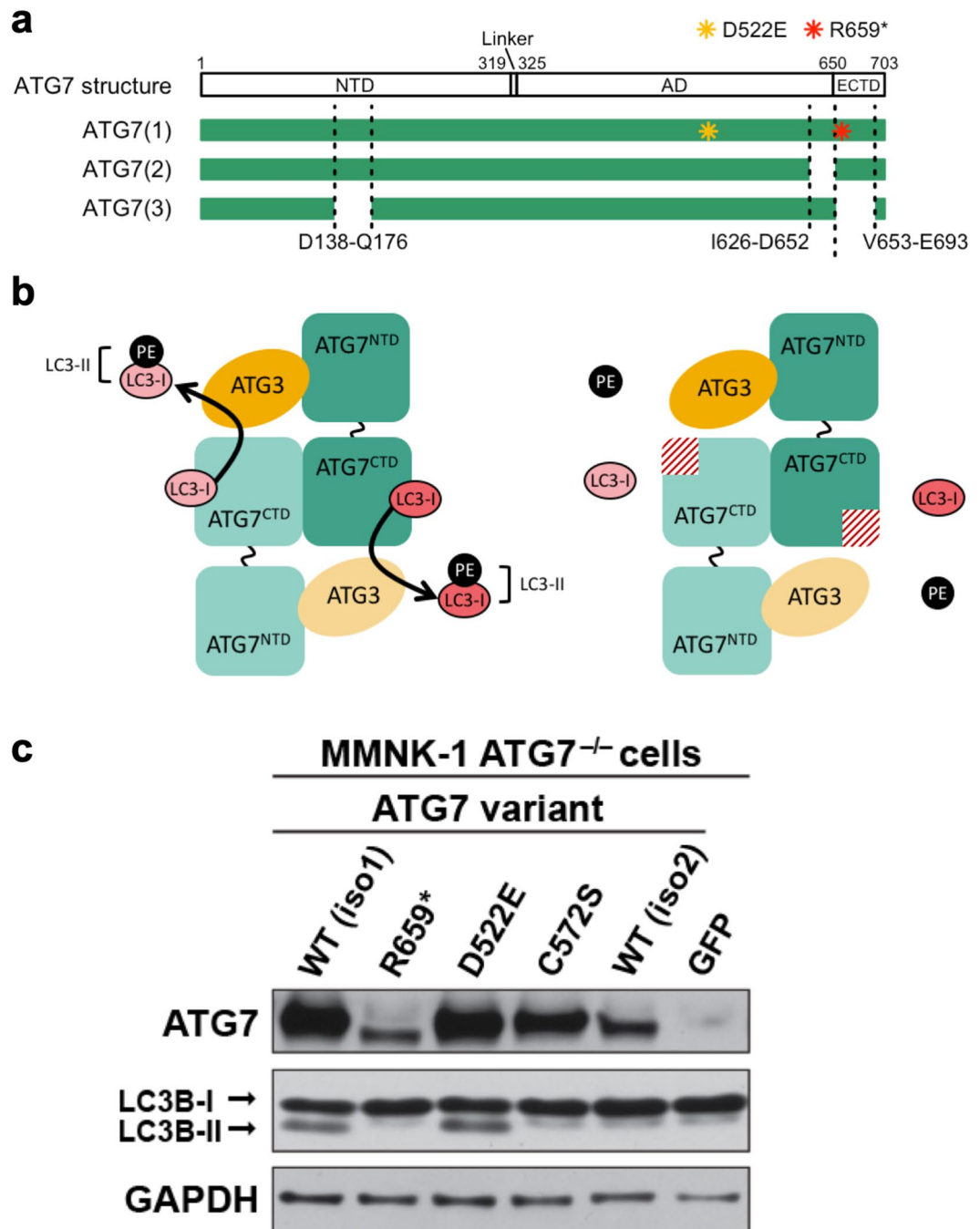
**Table 1.** Association analysis of rs146589465-G. Shown are associations for rs146589465-G in the Icelandic cohort that reached a threshold of *P*-value less than 5.0 × 10<sup>-2</sup>. The variant results in coding change p.D522E in *ATG7*. <sup>a</sup>A1 and A2 stand for the two alleles tested for the marker. Odds-ratio (OR) and population allele frequency (Freq. A1) are provided for allele A1.



**Figure 3.** Increased p62 expression in tumors of ATG7 variant (p.D522E) carriers. The microscopic images show sections of CCA from an rs146589465 (p.D522E) carrier (patient #17) and a non-carrier control (patient #20) stained with antibodies against ATG7, p62 and LC3B. For the LC3B staining, a higher magnification window is shown and the arrows point to LC3 dots that are indicative of autophagosomes. Scale bar represents 20  $\mu$ m and applies to all panels.

S21, Figs. 3 and S3). There was no difference in the expression of ATG7 when comparing tumors from carriers and non-carriers, suggesting that the variant does not alter the expression of the ATG7 protein. In addition, there was no difference in expression of the autophagy marker LC3 nor in the pattern of LC3 staining in the tissue samples. Interestingly, we observed significantly higher levels of cytoplasmic p62 protein in tumor samples from rs146589465-G carriers compared to those of non-carriers ( $P=0.003$ ). As noted, the accumulation of p62 is frequently used as an indicator for autophagy defects since this protein is degraded during normal autophagy activity<sup>36</sup>. For example, p62 levels are increased in hepatomas developed from *Atg7* knockout mice<sup>16</sup>.

**Protein structure implicates ATG7 R659\* as a putative loss-of-function mutation.** Using SWISS-MODEL we modeled and annotated the human ATG7 protein structure based on the yeast *Atg7* protein (Figs. 4a and S4)<sup>37</sup>. Human ATG7 has E1-like activity similar to yeast *Atg7*<sup>15</sup>. The yeast *Atg7* protein has been crystallized, revealing two conserved structural domains, namely the N-terminal domain (NTD) and the C-terminal domain (CTD)<sup>38,39</sup>. The NTD is responsible for Atg3 binding and the CTD is required for Atg8 (LC3/GABARAPs in mammals) recognition and binding. The C-terminal domain has two subdomains that include a homodimeric adenylation domain (AD) and an extreme C-terminal domain (ECTD). The human ATG7 p.R659 residue is in a highly conserved area of the ECTD and generates a stop codon leading to truncation of the ECTD



**Figure 4.** ATG7 p.R659\* is a loss-of-function mutation. (a) The predicted structure of ATG7 contains an N-terminal domain (NTD), linker region, adenylation domain (AD), and extreme C-terminal domain (ECTD). The p.R659\* (R659\*) and p.D522E (D522E) mutations are indicated in isoform 1 of the ATG7 protein. (b) Predicted functional impact of ATG7 p.R659\* mutation. Wildtype ATG7 and ATG3 conjugate ATG8 proteins such as LC3 with phosphatidylethanolamine (PE). We determined that the ATG7 R659\* mutant lacks PE conjugation activity to LC3 as a result of the ECTD truncation. (c) Expression of ATG7 WT isoform 1 (iso1) and D522E restored the lipidation of LC3B whereas expression of ATG7 R659\*, C572S, or WT isoform 2 (iso2) failed to lipidate LC3B from LC3B-I to LC3B-II. A Myc-tag was used to detect the expression of exogenous ATG7 from various vectors. A GFP expressing vector was used as a negative control. GAPDH was used as the loading control. Original blots are presented in Supplementary Fig. S9 online.

(see Supplementary Fig. S4 online). Therefore, we speculate that ATG7 p.R659 is a loss-of-function mutation and will be unable to bind LC3. The human ATG7 p.D522 residue is located on the exposed exterior loop of the AD in a less conserved area, which may affect protein interactions (see Supplementary Fig. S4 online).

**Evaluating the impact of the *ATG7* germline mutation and polymorphism on autophagy.** Human *ATG7* has three isoforms (Fig. 4A). Isoforms 1 and 2 are expressed in human tissues whereas isoform 3 is present at nearly undetectable levels<sup>40</sup>. *ATG7* isoform 2 lacks a part of the region required for LC3 binding and is unable to lipidate LC3<sup>40</sup>. Thus, only isoform 1 of *ATG7* has the capacity to lipidate LC3 during autophagosome formation. In yeast, the ECTD is essential for *Atg7* recognition and binding to *Atg8*. Along these lines, we conducted in vitro functional assays to determine whether the human *ATG7* variants identified in our study altered lipidation of LC3 (Fig. 4b)<sup>38,39</sup>. We hypothesized that the germline mutation from the family may have a more significant impact on specific autophagy functions than the population variant which was associated with a lower incidence of cholangiocarcinoma.

To study the functional impact of *ATG7* mutants, we engineered a human cholangiocyte cell line (MMNK-1) with an *ATG7* null background. The MMNK-1 cell line was derived from the human bile duct, thus representing the same primary tissue from which CCA arises. We applied CRISPR/Cas9 to generate a homozygous deletion in exon 2 of *ATG7* in MMNK-1 cells and isolated / expanded individual cells to screen for isogenic MMNK-1 *ATG7*<sup>-/-</sup> clones. We confirmed the complete loss of *ATG7* protein expression in MMNK-1 *ATG7*<sup>-/-</sup> cells (see Supplementary Fig. S5 online). LC3B lipidation was evaluated by western blotting where PE-bound LC3B-II. The lipidated form migrates faster than the non-lipidated LC3B-I. There was a notable reduction of LC3B-II in MMNK-1 *ATG7* CRISPR/Cas9 pools compared to the MMNK-1 cells, suggesting *ATG7* was successfully knocked out in majority of cells. Meanwhile, the isogenic MMNK-1 *ATG7*<sup>-/-</sup> cells derived from a single clone lacked a lower LC3B-II band, demonstrating complete loss of *ATG7* function (see Supplementary Fig. S5 online). The presence of the CRISPR-mediated homozygous deletion was confirmed with Sanger sequencing and targeted amplicon sequencing (see Supplementary Fig. S6 online).

To evaluate the function of the *ATG7* variants, isogenic MMNK-1 *ATG7*<sup>-/-</sup> cells were transfected with vectors expressing WT *ATG7* isoform 1, p.R659\* (familial germline mutation), p.D522E (the Icelandic population variant), p.C572S (a known inactivating mutation of *ATG7*), and wildtype *ATG7* isoform 2 (Methods). The expression level of *ATG7* R659\* was notably lower than the wildtype protein (Fig. 4c) despite similar levels of plasmid input and mRNA expression (see Supplementary Fig. S7 online). This result pointed to the truncating mutation impacting the stability of the *ATG7* protein. In comparison, the *ATG7* D522E protein was expressed at a similar level to the wildtype.

Subsequently, we assessed the activity of *ATG7* mutants by the level of lipidated LC3B. Transient expression of either wildtype *ATG7* isoform 1 or D522E enabled conversion of LC3B-I to LC3B-II in MMNK-1 *ATG7*<sup>-/-</sup> cells. In comparison, expression of R659\*, C572S, or the *ATG7* isoform 2 failed to convert LC3B-I to LC3B-II (Fig. 4c). Therefore, the R659\* mutation resulted in a loss-of-function of the *ATG7* lipidation activity.

To conduct long term functional assays of the *ATG7* variants and their impact on cholangiocytes, we established stable MMNK-1 *ATG7*<sup>-/-</sup> cell lines re-expressing *ATG7* WT and mutants by lentiviral transduction. There was substantially lower expression of *ATG7* protein in cell lines expressing the R659\* mutation (see Supplementary Fig. S8 online), which confirmed our previous observation from transient transfection in Fig. 4c.

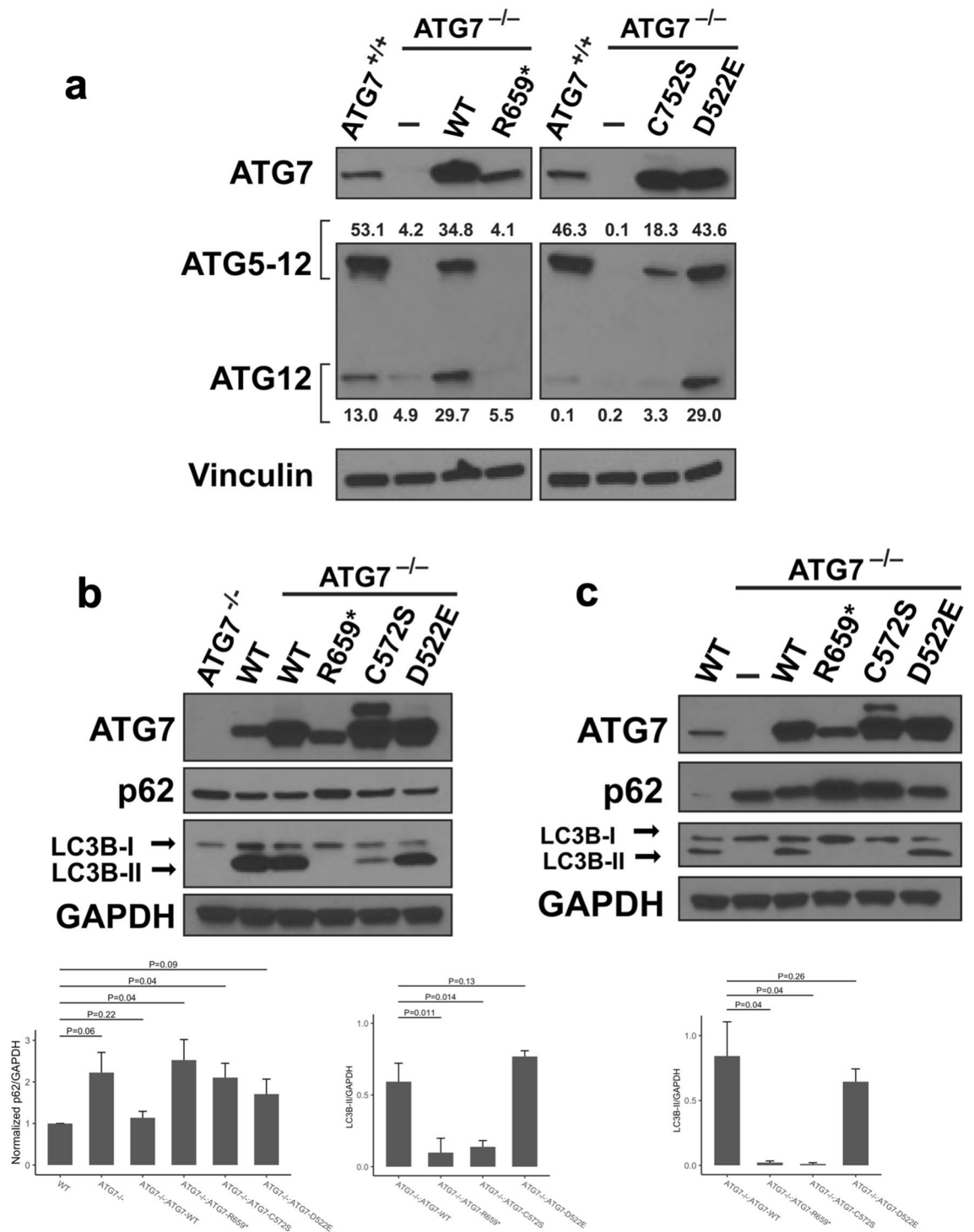
***ATG7* R659\* impairs autophagy by inhibiting the formation of the *ATG5-12* complex.** *ATG7* is required for two conjugation cascades that result in the formation of the *ATG5-12* complex and the subsequent lipidation of LC3/GABARAPs<sup>15,41,42</sup>. Our western blot studies showed that stable expression of the R659\* familial germline variant in an MMNK-1 *ATG7*<sup>-/-</sup> background failed to rescue the lipidation phenotype (see Supplementary Fig. S7 online). Therefore, we sought to determine whether the LC3 lipidation defect was the result of impaired *ATG5-12* formation via western blotting. We assayed *ATG12* using our stable cell line constructs in an *ATG7*<sup>-/-</sup> background<sup>43</sup>. Our results showed defects in the formation of the *ATG5-12* complex in *ATG7* null, R659\*- and C572S-expressing cell lines (Fig. 5a). Meanwhile, no abnormalities were noted for the *ATG7* WT and D522E variant when compared to parental MMNK-1 *ATG7*<sup>+/+</sup> cells. Similarly, we observed autophagy disruption by performing western blots against LC3B. When compared to *Atg7* wild-type rescue cells, a null rescue using R569\* or C572S resulted in impaired lipidation of LC3B-I to LC3B-II during basal and nutrient starvation conditions (Fig. 5b,c). No significant defect in LC3B-II lipidation was noted when D522E was used to rescue the *Atg7*-null cell line. Altogether, our results support a loss of function phenotype for the R659\* germline variant during autophagy.

***ATG7* variants lead to p62 accumulation.** To evaluate the autophagy response under metabolic stress, we assessed the level of ubiquitin-binding protein p62 in cholangiocytes expressing *ATG7* WT and mutants. The p62 protein is a key component for autophagic degradation of ubiquitinated proteins. Thus, p62 links the autophagy pathway and the ubiquitin-proteasome system for protein degradation<sup>44</sup>. We observed increased p62 levels in cholangiocytes expressing *ATG7* R659\* and C572S, and to a lesser extent, D522E during basal and starvation conditions (Fig. 5b,c). In addition, we observed that accumulation of p62 correlated with autophagy defects as evaluated by the impaired generation of LC3B-II in the null, R569\* and C572S genotypes under basal and starvation conditions (Fig. 5b,c). These results involved a comparison to the wild-type and WT *ATG7*, under basal conditions and metabolic stress under starvation conditions. Our data demonstrated that expression of cancer associated *ATG7* variants resulted in p62 accumulation in cholangiocyte cells under metabolic stress paralleling a tumorigenic environment.

## Discussion

In this study, we identified a novel germline mutation p.R659\* and a rare variant p.D522E in *ATG7* associated with CCA, based on a comprehensive genetic analysis of an extended pedigree of familial CCA and a population genetic association study, respectively. Our study of the CCA pedigree leveraged the analysis of haplotypes





**Figure 5.** Impaired ATG5-12 complex formation and elevated p62 levels in cholangiocytes expressing ATG7 R659\*. (a) Cholangiocytes expressing ATG7 R659\* unable to form ATG5-12 complex. The ATG7 R659\* and C572S variants fail to rescue the ATG5-12 complex formation defect observed in MMNK-1 ATG7<sup>-/-</sup> cholangiocytes. Protein lysates were obtained from cell lines under basal conditions. Free ATG12 and the ATG5-12 complex can be detected at 15 kDa and 55 kDa, respectively, on the western blot. (b) Top: Elevated p62 expression due to autophagy defects in cholangiocytes expressing ATG7 mutants. Cholangiocytes expressing ATG7 WT, R659\*, D522E, and C572S were cultured using normal growth media. Expression of ATG7, p62, LC3B, and GAPDH were detected on the western blot. Bottom left: Bar plot with y-axis showing normalized p62/GAPDH ratio and x-axis showing the distinct cell line genotypes. Normalized refers to the normalization of all genotypes in relation to the wild-type cell line. P-values are shown for statistical t-test comparisons. Bottom right: Bar plot with y-axis showing LC3B-II/GAPDH ratio and x-axis showing the distinct cell line genotypes. (c) Top: Cholangiocytes expressing ATG7 WT, R659\*, D522E, and C572S starved with HBSS for 3 h. Expression of ATG7, p62, LC3B, and GAPDH were detected on the western blot. Bottom: Bar plot with y-axis showing LC3B-II/GAPDH ratio and x-axis showing the distinct cell line genotypes. Original blots are presented in Supplementary Fig. S9 online. Quantifications reflect the mean of triplicates experiments for conditions and genotypes. P values are provided from t-tests performed.

shared among affected individuals in a family with CCA. We directly demonstrated that a somatic deletion of the *ATG7* wildtype allele in an affected tumor sample represented a second hit event.

*ATG7* plays an essential role in the autophagy pathway during autophagosome formation. Meanwhile, the role of autophagy in tumorigenesis is complicated, context-dependent and tissue-dependent<sup>45</sup>. Autophagy has been reported as tumor suppressive in some studies whereby autophagy deficiency led to tissue damage and genome instability. For example, heterozygous deletions of *BECN1* (*Atg6*), a member of the ATG gene family involved in autophagosome initiation, have been reported in human prostate, breast, and ovarian cancers, and *Becn1* deficiency led to spontaneous tumor development in mice<sup>46,47</sup>. Liver-specific autophagy deficiency due to *Atg7* deletion led to the development of liver tumors in mice<sup>17</sup>. In other studies, autophagy activity promoted tumor growth by suppressing metabolic stress and supporting cancer cell survival. For example, autophagy was required for the development of KRAS-driven lung adenocarcinomas in mice<sup>48</sup>. Previous studies have linked germline autophagy gene mutations to a variety of hereditary disorders. For example, inherited mutations in *ATG5* have been implicated in childhood ataxia, a neurodegenerative condition where balance and coordination are lost as well as leading to developmental delay<sup>49</sup>. Specific polymorphisms have been associated with an elevated risk of certain autoimmune conditions. Notably, *ATG5* and *ATG16L2* polymorphisms have been associated with systemic lupus erythematosus<sup>50,51</sup>, and *ATG16L1* polymorphisms have been linked to Crohn's disease<sup>52,53</sup>. However, the role of *ATG* family genes in cholangiocarcinoma remains unknown. Our study reports a germline mutation in *ATG7* strongly associated with CCA.

To study the functional impact of the *ATG7* variants in the context of cholangiocytes, we established a human *ATG7* null bile duct cell line. Our studies evaluating LC3 lipidation and Atg5-12 complex formation demonstrated that the germline mutation *ATG7* p.R659\* is a loss-of-function mutation during autophagy. We found that deficiency in autophagy due to *ATG7* p.R659\* resulted in the accumulation of p62 in cholangiocytes. This was notable under metabolic stress. Studies have demonstrated that accumulation of p62 results in the activation of antioxidant response, coupled with c-Myc and mTORC1 activation that results in metabolic reprogramming beneficial to cancer cells<sup>54,55</sup>.

The rare variant p.D522E identified from a population genetic association analysis did not display a noticeable effect on *ATG7* function, and *ATG7* p.D522E functioned like the wildtype protein with regard to LC3 lipidation and the conjugation cascades. This finding was not surprising because the population variant p.D522E resulted in a more modest change from aspartic acid to glutamic acid in a less conserved region compared to the p.R659\* mutation. However, p62 levels were elevated in the tumors of p.D522E carriers compared with control samples. The p62 protein is degraded by autophagy and thus its accumulation indicates impaired autophagy activity<sup>56</sup>. For example, notable p62 accumulation was observed in hepatocellular adenomas derived from liver-specific *Atg7* knockout mice<sup>16,57</sup>. Our findings suggest that the phenotypic effects of the p.D522E variant have less of an impact on autophagy. Therefore, the consequences of this polymorphism may be related to cumulative effects over a longer period of time, which could be seen in the tumors but is not detected in the cell culture model. In addition, the impact of p.D522E may occur through non-canonical autophagy pathways that are independent of LC3 lipidation or at later stages of canonical autophagy following lipidation. Further studies are needed to identify the precise cellular mechanisms affected by the p.D522E mutation.

In addition to its key role in autophagy, *ATG7* has been shown to bind p53 directly and regulate subsequent transcription of the critical cell-cycle inhibitor *CDKN1A*, which encodes the p21 protein<sup>58</sup>. Mouse embryonic stem cells deficient in *Atg7* fail to undergo cell cycle arrest, a defect not seen in *Atg5*- or *Atg6*-deficient embryos. The ability to directly bind to p53 and regulate cell cycle control is one potential mechanism by which the absence of *ATG7* could lead to the initiation of tumorigenesis. Further studies are needed to examine the effect of mutant *ATG7* on levels of reactive oxygen species (ROS) and DNA damage, and also to determine how increased levels of p62 may be linked to cholangiocarcinoma development through the p62-KEAP1-NRF2 axis. In autophagy-deficient mouse livers, p62 accumulation leads to the upregulation of cytoprotective genes resulting in toxicity and liver damage<sup>59</sup>. In future studies, we will investigate the precise mechanism of how *ATG7* loss-of-function contributes to tumorigenesis in cholangiocarcinoma.

## Materials and methods

**Clinical samples.** Our familial genetic study was conducted in compliance with the Helsinki Declaration, and the Institutional Review Board at Stanford University School of Medicine approved the study protocol (76,274). Samples were obtained with informed consent from the participants. Blood samples were obtained from eight siblings. We obtained matched tumor samples from the individuals affected with CCA (Supplementary Table S1 and Fig. 1). Genomic DNA extraction was performed using the Promega (Madison, WI) Maxwell 16 Blood DNA Purification Kit and 16 FFPE Plus LEV DNA Purification Kit for the germline blood samples and archival tumor samples, respectively.

**Exome sequencing and germline variant calling.** We prepared libraries with the KAPA Hyper Prep kit (Kapa Biosystems, Wilmington, MA) and exome capture was performed with the Nextera Rapid Capture Exome kit (Illumina, San Diego, CA). We sequenced the libraries on either the Illumina HiSeq 2500 system with 100 by 100-bp paired-end reads or the Illumina NextSeq system with 150 by 150-bp paired-end reads. The resulting sequence reads were aligned to human genome build GRCh37.1 with the BWA-MEM algorithm of the Burrows-Wheeler Aligner (BWA) v0.7.4<sup>60</sup>. WES and alignment metrics are available in Supplementary Table S2.

We used the Sentieon variant caller<sup>61</sup> (Mountain View, CA) v201711 to implement the GATK tools, following the GATK Best Practices for variant detection<sup>62</sup>. We generated a genome variant call format file for each sample using Sentieon Haplotyper and then performed joint variant calling on all eight samples with Sentieon GVCFTyper. We followed the GATK best practice recommendations for detecting SNVs and indels. Subsequently,

we filtered out low quality variants without a 'PASS' filter. The remaining variants were annotated with the following: Variant Effect Predictor v74<sup>63</sup> to assign general information including gene, amino acid, and mutation class; CADD v1.3<sup>64</sup> to generate a pathogenicity score for each variant; and ESP, ExAC and gnomAD<sup>65</sup> to determine the population allele frequency of each variant.

**Genome sequencing and haplotyping.** To determine haplotypes over extended genomic segments (i.e. Megabases) we used linked read whole genome sequencing. This sequencing approach enables one to use bar-coded short reads to obtain long-range genomic information from high molecular weight DNA<sup>66</sup>. With this long-range data, we performed germline SV detection and characterized the extended haplotypes across the entire genome for all eight siblings. For all of the pedigree germline samples, we prepared sequencing libraries with the Chromium Gel Bead and Library Kit and the Chromium Instrument (10X Genomics, Pleasanton, CA). These libraries were sequenced on the Illumina HiSeq X system with 150 by 150-bp paired-end reads. The resulting BCL files were converted to fastq files using Long Ranger (v2.1.2) 'mkfastq', then Long Ranger 'wgs' was run to align the reads to GRCh37.1, detect and phase SNVs/indels, and detect SVs. Linked read sequencing and alignment metrics are available in Supplementary Table S2, and phasing-haplotype metrics are available in Supplementary Table S3.

For the detection of somatic SNVs and indels, we performed high depth WES on three CCA samples in combination with the results from the matched normal blood samples in the germline analysis (above). Sequencing libraries for two of the tumor samples (patient III:3 and III:8) were generated using the KAPA Hyper Prep kit (Kapa Biosystems). The other tumor sample (patient III:4) had lower quality DNA and was prepared using the GemCode Library Kit (10X Genomics) to retain long-range information for improved read alignment. Exome capture was performed with the Nextera Rapid Capture Exome kit (Illumina). We sequenced the libraries on the Illumina HiSeq system with 100 by 100-bp paired-end reads or the Illumina NextSeq system with 150 by 150-bp paired-end reads. The sequence reads were aligned to human genome build GRCh37.1 with the BWA-MEM algorithm of BWA v0.7.4<sup>60</sup>. For the GemCode library, the BCL files were converted to fastq files using Long Ranger (v2.1.2) 'mkfastq', then Long Ranger (v2.1.2) 'wgs' was run to align the reads to GRCh37.1.

The mean on-target coverage for the tumor WES data ranged from 88 to 178×; sequencing and alignment metrics are available in Supplementary Table S2. We used the Sentieon v201711 TNsnv tool (similar to the MuTect algorithm) and TNhaplotyper tool (similar to the MuTect2 algorithm) to detect somatic SNVs and small indels using the WES data of each tumor/normal pair and retained only those variants with a 'PASS' filter. We annotated the variants with the same methodology as in the 'Germline analysis' (above), and also applied the same filters: the variant had to have the potential to affect protein function and the variant had to occur at low frequency (i.e. less than 1%) in population frequency data. For patient III:4, we used the WES data to identify CNVs using the CNVkit tool and determined copy number changes<sup>67</sup>.

We performed WGS on tumor-normal pairs from patients III:3 and III:8. Sequencing libraries were generated using the KAPA Hyper Prep kit (Kapa Biosystems). The libraries for patient III:3 were sequenced on the Illumina HiSeq 4000 system with 150 by 150-bp paired-end reads and the libraries for patient III:8 were sequenced on the Illumina HiSeq X system with 150 by 150-bp paired-end reads. The resulting sequence reads were aligned to human genome build GRCh37.1 with the BWA-MEM algorithm of BWA v0.7.4<sup>60</sup>. WGS and alignment metrics are available in Supplementary Table S2. To identify candidate genes where the second (i.e. 'wildtype') allele could have been lost due a genomic deletion or gene conversion, we ran the BICseq2<sup>68</sup> CNV detection tool on these tumor-normal pairs.

We used linked read WGS to retain long-range genomic information from one of the tumor samples (patient III:8). We prepared the sequencing library for the tumor sample using the Chromium Library Kit (10X Genomics). The library was sequenced on the Illumina HiSeq X system with 150 by 150-bp paired-end reads. The resulting BCL files were converted to fastq files using Long Ranger (v2.1.2) 'mkfastq', then Long Ranger (v2.1.2) 'wgs' was run to align the reads to GRCh37.1, detect and phase SNVs/indels, and detect SVs.

We determined whether each candidate causal mutation was in *cis* or in *trans* with somatic deletion events in the tumor. For this haplotype analysis, we used linked read data to generate digital karyotypes as described in Bell et al.<sup>34</sup>. This method generates extended haplotypes from Mb segments, by combining adjacent phase blocks extrapolated from allelic imbalances. The Mb haplotypes are quantitatively determined from the differences in the linked read barcode counts among haplotypes. Using this process, we determined the alleles that belonged to each haplotype in regions with genomic deletions.

**Population genetic study of CCA.** Individuals affected with CCA were identified through the ICR and samples were obtained with informed consent from the participants. All sample identifiers were encrypted in accordance with the regulations of the Icelandic Data Protection Authority. Approval for the study was granted by the Icelandic National Bioethics Committee (ref. 00/097) and the Icelandic Data Protection Authority (ref. 2001020223).

We identified all of the known variants occurring in *ATG7* for the Icelandic population as previously described<sup>35</sup>. A total of 31.6 million SNVs and short indels that met quality control criteria were identified in the genomes of 8453 sequenced Icelanders. These variants were then imputed into 150,656 Icelanders genotyped with Illumina SNP chips, and their genotypes were phased using long-range phasing<sup>69,70</sup>. To increase the sample size and power for genetic association analysis, we used genealogical deduction of carrier status for 294,212 relatives lacking array-based genotypes. Subsequently, *ATG7* variant association testing for case-control analysis was performed using logistic regression. The quality of the imputation was evaluated by comparing imputed genotypes to genotypes obtained by direct genotyping. Individual *ATG7* genotyping was performed by applying the Centaurus (Nanogen) platform.

**Tumor immunohistochemistry.** Immunohistochemistry was performed on 3  $\mu\text{m}$  sections from paraffin-embedded tumors. Following deparaffinization, samples were rehydrated and subjected to heat-induced epitope retrieval. Tris/EDTA buffer pH 9 was used for the ATG7 (clone EP1759Y, Millipore (Burlington, MA) 04–1055; 1:1000), LC3B (clone D11, Cell Signaling Technology (Danvers, MA) 3868; 1:250) and p62 (Enzo (New York, NY) PW9860; 1:100) antibodies in a 98.2 °C water bath (5/20/20). Endogenous peroxidase activity was blocked with 3% hydrogen peroxide. After incubation with the respective primary antibodies for 30 min at room temperature, EnVision FLEX Kit (DAKO, Glostrup, Denmark) was used for detection. Microscope images of stained sections were scored in a blinded manner. A chi-square test was performed to assess whether staining patterns significantly differed between carrier and control samples.

**Cell lines.** We used both human and mouse cell lines for these experiments. The immortalized human cholangiocyte cell line MMNK-1, and mouse embryonic fibroblasts (MEF) *Atg7*<sup>-/-</sup> and MEF *Atg7*<sup>+/+</sup>, were obtained from RIKEN BioResource Center Cell Bank (Wako, Saitama, Japan). Cells were maintained in DMEM supplemented with 10% fetal bovine serum at 37 °C and 5% CO<sub>2</sub>. For characterizing these cells, we extracted genomic DNA and total RNA using the Maxwell 16 Cell LEV DNA Purification Kit and the Maxwell 16 Cell LEV Total RNA Purification Kit (Promega), respectively.

**Generation, screening, and genotyping of isogenic MMNK-1 ATG7<sup>-/-</sup> cell lines.** Using CRISPR-Cas9, exon 2 of *ATG7* was targeted for knockout in the MMNK-1 cell population (ZeClinics, Barcelona, Spain). Guide RNAs targeting exon 2 of *ATG7* were cloned into the PX458 and PX459 vectors. The guide RNA sequences are listed below.

*ATG7* sgRNA22\_Top: 5'- CACCGAACTGCAGTTTAgAGAGTCC -3'.

*ATG7* sgRNA22\_Bottom: 5'- AAACGGACTCTCTAAACTGCAGTTC -3'.

*ATG7* sgRNA119\_Top: 5'- CACCGAAGCTGAACGAGTATCGGC -3'.

*ATG7* sgRNA119\_Bottom: 5'- AAACGCCGATACTCGTTCAGCTTC -3'.

MMNK-1 cells were transfected with both sgRNA vectors and then selected with puromycin. We sorted single MMNK-1 cells into 96 wells to generate isogenic MMNK-1 *ATG7*<sup>-/-</sup> cell clones.

To screen for isogenic MMNK-1 *ATG7*<sup>-/-</sup> clones, we examined the expression of ATG7 protein using western blotting and selected the individual clones with complete loss of ATG7 expression for further expansion.

To genotype the MMNK-1 *ATG7*<sup>-/-</sup> clones and confirm the presence of deletion in the exon 2 of *ATG7*, we applied both sanger sequencing and Illumina short read sequencing to sequence the area targeted by CRISPR/Cas9 knockout following the standard procedure as described previously<sup>71</sup>. The primers used to generate the PCR amplicon for sequencing are list below:

*ATG7*\_exon\_F: GTTGTGTTTCAAGGTAGCCTGT.

*ATG7*\_exon\_R: TCTCTTCCACAGCAATGCT.

**Human *ATG7* variant plasmids for transient transfection experiments.** The plasmid pCMV-myc-Atg7(2) expressing *ATG7* isoform 2 was a gift from Toren Finkel (Addgene plasmid #24921)<sup>72</sup>. The plasmid pCMV-myc-Atg7(1) expressing *ATG7* isoform 1 was derived from pCMV-myc-Atg7(2) using the Q5 site-directed mutagenesis kit (NEB, Ipswich, MA). Subsequently, we used pCMV-myc-Atg7(1) to introduce the *ATG7* mutations p.R659\*, p.D522E and p.C572S using the Q5 site-directed mutagenesis kit. After subcloning, all of the *ATG7* mutations in the vectors were confirmed with Sanger sequencing. The pCMV3-C-GFPspark vector was used as the control (Sino Biological, Beijing, China). The primers are listed below:

*ATG7* p.R659\*\_F: 5'- GTTCTTGATCAATATGAATGAGAAGGATTTAAC -3'.

*ATG7* p.R659\*\_R: 5'- TTTGGAAGAACAAGCTGTACATTTG -3'.

*ATG7* p.D522E\_F: 5'- AGCTGGGGAGTTGTGTCCA -3'.

*ATG7* p.D522E\_R: 5'- CCTTGCTGCTTTGGTTTCTTCA -3'.

*ATG7* p.C572S\_F: 5'- CTTGGACCAGCAGTCCACTGTGAGTCG -3'.

*ATG7* p.C572S\_R: 5'- GTCCGGTCTCTGGTTGAATCTCCTGG -3'.

**Transient transfection of *ATG7* with mutations.** Human MMNK-1 *ATG7*<sup>-/-</sup> or mouse MEF *Atg7*<sup>-/-</sup> cells were transfected with 2.5  $\mu\text{g}$  of plasmid expressing WT *ATG7* (isoform 1), *ATG7* p.R659\*, *ATG7* p.D522E, *ATG7* p.C572S or GFP control, using Lipofectamine 2000. Cells were starved in EBSS for 3 h and harvested for protein extraction 24 h post-transfection.

**Western blotting.** Cell lysates (20  $\mu\text{g}$ ) were separated on 4–20% precast polyacrylamide gels (Bio-Rad, Hercules, CA) and were transferred to nitrocellulose membranes. The following antibodies were used for immunoblotting: *ATG7* (D12B11, Cell Signaling), *ATG12* (D88H11, Cell Signaling), LC3B (D11, Cell Signaling), p62/SQSTM1 (D-3, Santa Cruz Biotechnology (Dallas, TX)), Myc-Tag (9B11, Cell Signaling), Vinculin (E1E9V, Cell Signaling), and GAPDH (14C10, Cell Signaling).

For quantifications, each western blot experiment was performed as three independent biological replicates for all conditions and genotypes. Western blot films were exported to Adobe Photoshop CS3 and the images were inverted to black and white followed by inversion of the image for band measurements. The box tool was used to obtain the mean gray value of the film background and each individual band. Each western blot band was normalized to a loading internal control for quantification (e.g. GAPDH). Student's *t*-test statistical analysis was used to compare all genotypes and conditions tested using the mean from triplicate measurements.

**Quantification of *ATG7* plasmid input and gene expression.** We determined the amount of plasmid maintained in the cell lines after transient transfection using droplet digital PCR (ddPCR). The ddPCR was performed on the Bio-Rad QX200 ddPCR system following the manufacturer instructions. Briefly, each 22  $\mu$ L ddPCR reaction used 1X EvaGreen ddPCR super-mix, 100 nM of primers and  $10^{-5}$  ng of plasmid. The PCR master mix was partitioned and plated per Bio-Rad's QX200 droplet generation protocol. After thermal cycling, the plate was transferred and read by the Bio-Rad QX200 Droplet Reader. Results were analyzed with the Quantasoft program (Bio-Rad).

The expression of mRNA transcripts from the cell lines was determined using a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA) with SYBR Green master mixes (Bio-Rad) following the manufacturer instructions. The *RPP30* gene was used as a housekeeping control. The primers used for *ATG7* ddPCR and cDNA amplification are listed below:

*ATG7\_F*: 5'-AAGCCATGATGTCGTCTTCC-3'

*ATG7\_R*: 5'-TCCTTGCTGCTTTGGTTTCT-3'.

**Lentiviral constructs of the human *ATG7* variants.** Constructs were generated using the BP/LR Gateway cloning system (Invitrogen, Cambridge, UK). Briefly, the *ATG7* constructs described above (WT, R659\*, D522E, and C572S) were cloned from pENTR21 into the final vector pLenti-CMV-Blast-Destiny. After antibiotic selection, single colonies were used to inoculate liquid LB for subsequent use in plasmid extraction. The *ATG7* variants generated above were used to produce lentiviruses, which were subsequently transduced into MMNK-1 *ATG7*<sup>-/-</sup> cells. Briefly, the lentiviral vectors were co-transfected with plasmids psPAX2 and pMD2.G into HEK293 cells at ~80–90% confluency. The lentivirus particles were harvested from cell media 48 h post-transfection and precipitated using PEG-it virus Precipitation Solution (Abcam, Cambridge, UK). Transduced MMNK-1 *ATG7*<sup>-/-</sup> cells were selected by blasticidin.

### Data availability

All sequence data for this study has been submitted to dbGaP as fastq or bam files at the accession phs001593.v1.p1.

Received: 29 November 2021; Accepted: 25 May 2022

Published online: 20 June 2022

### References

- Levine, B. & Kroemer, G. Biological functions of autophagy genes: A disease perspective. *Cell* **176**, 11–42. <https://doi.org/10.1016/j.cell.2018.09.048> (2019).
- Dikic, I. & Elazar, Z. Mechanism and medical implications of mammalian autophagy. *Nat. Rev. Mol. Cell Biol.* **19**, 349–364. <https://doi.org/10.1038/s41580-018-0003-4> (2018).
- Santoni, M. *et al.* Autophagic gene polymorphisms in liquid biopsies and outcome of patients with metastatic clear cell renal cell carcinoma. *Anticancer Res.* **38**, 5773–5782. <https://doi.org/10.21873/anticancer.12916> (2018).
- Maitra, A. & Hruban, R. H. Pancreatic cancer. *Annu. Rev. Pathol. Mech. Dis.* **3**, 157–188. <https://doi.org/10.1146/annurev.pathol.3.121806.154305> (2008).
- Shaib, Y. H., Davila, J. A., McGlynn, K. & El-Serag, H. B. Rising incidence of intrahepatic cholangiocarcinoma in the United States: A true increase?. *J. Hepatol.* **40**, 472–477. <https://doi.org/10.1016/j.jhep.2003.11.030> (2004).
- DeOliveira, M. L. *et al.* Cholangiocarcinoma: Thirty-one-year experience with 564 patients at a single institution. *Ann. Surg.* **245**, 755–762. <https://doi.org/10.1097/01.sla.0000251366.62632.d3> (2007).
- Farshidfar, F. *et al.* Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep.* **19**, 2878–2880. <https://doi.org/10.1016/j.celrep.2017.06.008> (2017).
- Jusakul, A. *et al.* Whole-genome and epigenomic landscapes of etiologically distinct subtypes of Cholangiocarcinoma. *Cancer Discov.* **7**, 1116–1135. <https://doi.org/10.1158/2159-8290.CD-17-0368> (2017).
- Wardell, C. P. *et al.* Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *J. Hepatol.* **68**, 959–969. <https://doi.org/10.1016/j.jhep.2018.01.009> (2018).
- Xue, R. *et al.* Genomic and transcriptomic profiling of combined hepatocellular and intrahepatic cholangiocarcinoma reveals distinct molecular subtypes. *Cancer Cell* **35**, 932–947. <https://doi.org/10.1016/j.ccell.2019.04.007> (2019).
- Yorimitsu, T. & Klionsky, D. J. Autophagy: Molecular machinery for self-eating. *Cell. Death Differ.* **12**(Suppl 2), 1542–1552. <https://doi.org/10.1038/sj.cdd.4401765> (2005).
- Kabeya, Y. *et al.* LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J. Cell Sci.* **117**, 2805–2812. <https://doi.org/10.1242/jcs.01131> (2004).
- Johansen, T. & Lamark, T. Selective autophagy: ATG8 family proteins, LIR motifs and cargo receptors. *J. Mol. Biol.* **432**, 80–103. <https://doi.org/10.1016/j.jmb.2019.07.016> (2020).
- Tanida, I. *et al.* Apg7p/Cvt2p: A novel protein-activating enzyme essential for autophagy. *Mol. Biol. Cell* **10**, 1367–1379. <https://doi.org/10.1091/mbc.10.5.1367> (1999).
- Tanida, I., Tanida-Miyake, E., Ueno, T. & Kominami, E. The human homolog of *Saccharomyces cerevisiae* Apg7p is a Protein-activating enzyme for multiple substrates including human Apg12p, GATE-16, GABARAP, and MAP-LC3. *J. Biol. Chem.* **276**, 1701–1706. <https://doi.org/10.1074/jbc.C000752200> (2001).
- Takamura, A. *et al.* Autophagy-deficient mice develop multiple liver tumors. *Genes Dev.* **25**, 795–800. <https://doi.org/10.1101/gad.2016211> (2011).
- Lee, Y. A. *et al.* Autophagy is a gatekeeper of hepatic differentiation and carcinogenesis by controlling the degradation of Yap. *Nat. Commun.* **9**, 4962–4962. <https://doi.org/10.1038/s41467-018-07338-z> (2018).
- Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35. <https://doi.org/10.1038/ng.499> (2010).
- Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755. <https://doi.org/10.1038/nrg3031> (2011).
- Aarnio, M. *et al.* Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int. J. Cancer* **81**, 214–218. [https://doi.org/10.1002/\(Sici\)1097-0215\(19990412\)81:2%3c214::Aid-Ijc8%3e3.3.Co;2-C](https://doi.org/10.1002/(Sici)1097-0215(19990412)81:2%3c214::Aid-Ijc8%3e3.3.Co;2-C) (1999).

21. Klebe, S. *et al.* BAP1 hereditary cancer predisposition syndrome: a case report and review of literature. *Biomark. Res.* **3**, 14. <https://doi.org/10.1186/s40364-015-0040-5> (2015).
22. Pilarski, R. *et al.* Expanding the clinical phenotype of hereditary BAP1 cancer predisposition syndrome, reporting three new cases. *Genes Chromosomes Cancer* **53**, 177–182. <https://doi.org/10.1002/gcc.22129> (2014).
23. Nakanuma, Y. & Sato, Y. Hilar cholangiocarcinoma is pathologically similar to pancreatic duct adenocarcinoma: Suggestions of similar background and development. *J. Hepatobiliary Pancreat. Sci.* **21**, 441–447. <https://doi.org/10.1002/jhbp.70> (2014).
24. Thompson, D., Easton, D. F. & Breast Cancer Linkage, C. Cancer Incidence in BRCA1 mutation carriers. *J. Natl. Cancer Inst.* **94**, 1358–1365. <https://doi.org/10.1093/jnci/94.18.1358> (2002).
25. Hahn, S. A. *et al.* BRCA2 germline mutations in familial pancreatic carcinoma. *Gastroenterology* **124**, A548–A548. [https://doi.org/10.1016/S0016-5085\(03\)82772-5](https://doi.org/10.1016/S0016-5085(03)82772-5) (2003).
26. Jones, S. *et al.* Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* **324**, 217. <https://doi.org/10.1126/science.1171202> (2009).
27. Roberts, N. J. *et al.* ATM mutations in patients with hereditary pancreatic cancer. *Cancer Discov.* **2**, 41–46. <https://doi.org/10.1158/2159-8290.Cd-11-0194> (2012).
28. Su, G. H. *et al.* Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. *Am. J. Pathol.* **154**, 1835–1840. [https://doi.org/10.1016/S0002-9440\(10\)65440-5](https://doi.org/10.1016/S0002-9440(10)65440-5) (1999).
29. Bartsch, D. K. *et al.* CDKN2A germline mutations in familial pancreatic cancer. *Ann. Surg.* **236**, 730–737. <https://doi.org/10.1097/01.Sla.0000036393.89509.4e> (2002).
30. Rafnar, T. *et al.* Mutations in BRIP1 confer high risk of ovarian cancer. *Nat. Genet.* **43**, 1104–1107. <https://doi.org/10.1038/ng.955> (2011).
31. Grant, R. C. *et al.* Exome-wide association study of pancreatic cancer risk. *Gastroenterology* **154**, 719–722 e713. <https://doi.org/10.1053/j.gastro.2017.10.015> (2018).
32. Klein, A. P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat. Commun.* **9**, 556. <https://doi.org/10.1038/s41467-018-02942-5> (2018).
33. Tamura, K. *et al.* Mutations in the pancreatic secretory enzymes CPA1 and CPB1 are associated with pancreatic cancer. *Proc. Natl. Acad. Sci. U S A* **115**, 4767–4772. <https://doi.org/10.1073/pnas.1720588115> (2018).
34. Bell, J. M. *et al.* Chromosome-scale mega-haplotypes enable digital karyotyping of cancer aneuploidy. *Nucleic Acids Res.* **45**, e162. <https://doi.org/10.1093/nar/gkx712> (2017).
35. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444. <https://doi.org/10.1038/ng.3247> (2015).
36. Carlsson, S. R. & Simonsen, A. Membrane dynamics in autophagosome biogenesis. *J. Cell. Sci.* **128**, 193–205. <https://doi.org/10.1242/jcs.141036> (2015).
37. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303. <https://doi.org/10.1093/nar/gky427> (2018).
38. Noda, N. N. *et al.* Structural basis of Atg8 activation by a homodimeric E1, Atg7. *Mol. Cell* **44**, 462–475. <https://doi.org/10.1016/j.molcel.2011.08.035> (2011).
39. Taherbhoy, A. M. *et al.* Atg8 transfer from Atg7 to Atg3: A distinctive E1–E2 architecture and mechanism in the autophagy pathway. *Mol. Cell* **44**, 451–461. <https://doi.org/10.1016/j.molcel.2011.08.034> (2011).
40. Ogmundsdottir, M. H. *et al.* A short isoform of ATG7 fails to lipidate LC3/GABARAP. *Sci. Rep.* **8**, 14391. <https://doi.org/10.1038/s41598-018-32694-7> (2018).
41. Mizushima, N. *et al.* A protein conjugation system essential for autophagy. *Nature* **395**, 395–398. <https://doi.org/10.1038/26506> (1998).
42. Xiong, J. Atg7 in development and disease: Panacea or Pandora's Box?. *Protein Cell* **6**, 722–734. <https://doi.org/10.1007/s13238-015-0195-8> (2015).
43. Nitta, A. *et al.* Blocking LC3 lipidation and ATG12 conjugation reactions by ATG7 mutant protein containing C572S. *Biochem. Biophys. Res. Commun.* **508**, 521–526. <https://doi.org/10.1016/j.bbrc.2018.11.158> (2019).
44. Liu, W. J. *et al.* p62 links the autophagy pathway and the ubiquitin-proteasome system upon ubiquitinated protein degradation. *Cell Mol. Biol. Lett.* **21**, 29. <https://doi.org/10.1186/s11658-016-0031-z> (2016).
45. White, E. The role for autophagy in cancer. *J. Clin. Invest.* **125**, 42–46. <https://doi.org/10.1172/JCI73941> (2015).
46. Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl. Acad. Sci. U S A* **100**, 15077–15082. <https://doi.org/10.1073/pnas.2436255100> (2003).
47. Qu, X. *et al.* Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *J. Clin. Invest.* **112**, 1809–1820. <https://doi.org/10.1172/JCI20039> (2003).
48. Guo, J. Y. & White, E. Autophagy is required for mitochondrial function, lipid metabolism, growth, and fate of KRAS(G12D)-driven lung tumors. *Autophagy* **9**, 1636–1638. <https://doi.org/10.4161/auto.26123> (2013).
49. Kim, M. *et al.* Mutation in ATG5 reduces autophagy and leads to ataxia with developmental delay. *Elife* **5**, 1. doi:<https://doi.org/10.7554/eLife.12245> (2016).
50. Zhou, X. J. *et al.* Genetic association of PRDM1-ATG5 intergenic region and autophagy with systemic lupus erythematosus in a Chinese population. *Ann. Rheum. Dis.* **70**, 1330–1337. <https://doi.org/10.1136/ard.2010.140111> (2011).
51. Molineros, J. E. *et al.* Confirmation of five novel susceptibility loci for systemic lupus erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci. *Hum. Mol. Genet.* **26**, 1205–1216. <https://doi.org/10.1093/hmg/ddx026> (2017).
52. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39**, 207–211. <https://doi.org/10.1038/ng1954> (2007).
53. Glas, J. *et al.* The ATG16L1 gene variants rs2241879 and rs2241880 (T300A) are strongly associated with susceptibility to Crohn's disease in the German population. *Am. J. Gastroenterol.* **103**, 682–691. <https://doi.org/10.1111/j.1572-0241.2007.01694.x> (2008).
54. Saito, T. *et al.* p62/Sqstm1 promotes malignancy of HCV-positive hepatocellular carcinoma through Nrf2-dependent metabolic reprogramming. *Nat. Commun.* **7**, 12030–12030. <https://doi.org/10.1038/ncomms12030> (2016).
55. Umemura, A. *et al.* p62, Upregulated during preneoplasia, induces hepatocellular carcinogenesis by maintaining survival of stressed HCC-Initiating cells. *Cancer Cell* **29**, 935–948. <https://doi.org/10.1016/j.ccell.2016.04.006> (2016).
56. Bjorkoy, G. *et al.* Monitoring autophagic degradation of p62/SQSTM1. *Methods Enzymol.* **452**, 181–197. [https://doi.org/10.1016/S0076-6879\(08\)03612-4](https://doi.org/10.1016/S0076-6879(08)03612-4) (2009).
57. Inami, Y. *et al.* Persistent activation of Nrf2 through p62 in hepatocellular carcinoma cells. *J. Cell Biol.* **193**, 275–284. <https://doi.org/10.1083/jcb.201102031> (2011).
58. Lee, I. H. *et al.* Atg7 modulates p53 activity to regulate cell cycle and survival during metabolic stress. *Science* **336**, 225–228. <https://doi.org/10.1126/science.1218395> (2012).
59. Komatsu, M. *et al.* The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat. Cell Biol.* **12**, 213–223. <https://doi.org/10.1038/ncb2021> (2010).
60. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
61. Freed, D. N., Aldana, R., Weber, J. A. & Edwards, J. S. The Sentieon Genomics Tools—A fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv* <https://doi.org/10.1101/115717> (2017).

62. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. <https://doi.org/10.1101/gr.107524.110> (2010).
63. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**. <https://doi.org/10.1186/s13059-016-0974-4> (2016).
64. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894. <https://doi.org/10.1093/nar/gky1016> (2019).
65. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–+, doi:<https://doi.org/10.1038/nature19057> (2016).
66. Zheng, G. X. *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **34**, 303–311. <https://doi.org/10.1038/nbt.3432> (2016).
67. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *Plos Comput. Biol.* **12**, 1. <https://doi.org/10.1371/journal.pcbi.1004873> (2016).
68. Xi, R. B., Lee, S., Xia, Y. C., Kim, T. M. & Park, P. J. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. *Nucleic Acids Res.* **44**, 6274–6286. <https://doi.org/10.1093/nar/gkw491> (2016).
69. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075. <https://doi.org/10.1038/ng.216> (2008).
70. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103. <https://doi.org/10.1038/nature09525> (2010).
71. Chen, J. *et al.* Recurrent DICER1 hotspot mutations in endometrial tumours and their impact on microRNA biogenesis. *J. Pathol.* **237**, 215–225. <https://doi.org/10.1002/path.4569> (2015).
72. Lee, I. H. *et al.* A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. *Proc. Natl. Acad. Sci. USA* **105**, 3374–3379. <https://doi.org/10.1073/pnas.0712145105> (2008).

## Acknowledgements

The authors recognize and appreciate the patients and families who contributed to the current study. We acknowledge the Icelandic Cancer Registry for assistance in the ascertainment of the Icelandic cancer patients. We thank deCODE genetics for access to data and facilities, assistance with data analysis and helpful discussions. This work was supported by the National Institutes of Health [P01HG000205 to SUG and HPJ, 1U01CA15192001-A1 to HPJ, 1U01CA176299 to HPJ, HG006137-07 to HPJ, R01 CA116468NIH to DAJ, 5K08CA166512 to LDN]; Intermountain Healthcare to SUG, JC and HPJ; a Research Scholar Grant from the American Cancer Society [RSG-13-297-01-TBG to JC and HPJ]; Clayville Foundation to HPJ; Gastric Cancer Foundation to HPJ and LDN; the Samuel Waxman Cancer Research Foundation to DAJ; Oklahoma Center for Adult Stem Cell Research (OCASCR) to DAJ; Oklahoma Medical Research Foundation (OMRF) to DAJ; the Conquer Cancer Foundation (Young Investigator Award) to LDN; the Carl Kawaja Foundation to LDN; the Research Fund of Iceland [130230-0529 to ES and MHO, 184861-052 to ES, 184727-051 to MHO]; and a grant from the Icelandic Cancer Society Research Fund to MHO.

## Author contributions

L.D.N. and D.S.H., with assistance from R.R. and G.F., collected the patient data and samples for the family with C.C.A.. B.T.L. conducted the whole genome, whole exome and linked read whole genome sequencing assays. S.U.G. and H.P.J. analyzed the sequencing data and performed the genetic analysis; the mega-haplotyping analysis was performed by J.M.B.. Statistical analysis of Icelandic population data was performed by M.H.O. Selection and preparation of paraffin samples, staining, imaging and scoring of patient samples was performed by J.G.J., S.K., and M.H.O. The Icelandic population study and FFPE staining experiments were designed and results interpreted by M.H.O. and E.S.. Cell line experiments were designed and performed by J.C. and C.A. R.C.D., I.T.S., and D.A.J. contributed to the writing of the manuscript and the scientific interpretation regarding the genetics of *ATG7*. S.U.G., J.C., M.H.O., E.S., J.C., D.A.J., H.P.J., and L.D.N. wrote the manuscript. L.D.N. and H.P.J. designed, supervised, and coordinated all aspects of the research. All authors revised, read, and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13569-4>.

**Correspondence** and requests for materials should be addressed to H.P.J. or L.D.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022