Material & Methods: (EXTENDED ONLINE METHODS)

Genomic analysis: Low-density lipoprotein receptor (LDLR) genes from representative species across vertebrate phylogeny were identified and examined for orthology to human LDLR, using reciprocal cross alignments and synteny verification. We identified LDLR annotations in NCBI with sufficient synteny information for orthology analysis in 12 bird species, including five songbirds (*Taeniopygia guttata*, *Pseudopodoces humilis*, *Corvus moneduloides*, *Camarhynchus parvulus*, *Catharus ustulatus*), two parrots (*Melopsittacus undulatus*, *Strigops habroptila*), one eagle (*Aquila chrysaetos chrysaetos*), two Galliformes (*Coturnix japonica*, *Gallus gallus*) and two ratites (*Apteryx mantelli mantelli*, *Apteryx rowi*), noting that the predictions listed as low-density lipoprotein receptor-like in one owl (*Tyto alba*) and one songbird (*Ficedula albicollis*) are misannotations (most likely a paralog of LDLR-related protein 2 - LRP-2 - in both species) based on synteny and sequence similarity. In chicken, a partial LDLR prediction is present in an unplaced scaffold of the current assembly (Gene ID: 395103 in galGal6), flanked by gaps. BLAST searches of the corresponding chicken Pacbio p-read dataset using the full cloned chicken LDLR mRNA (NM_204452) and the predicted transcripts for syntenic genes from other avian species identified p-reads containing subsets of the exons from LDLR and syntenic genes. Based on these, we were able to reconstruct a putative synteny in chicken (https://osf.io/hdrvy/), which was the same as in other birds (Fig. 1b). A similar strategy identified LDLR and syntenic genes in Pacbio p-reads of a hummingbird (*Calypte anna*) (20). For non-avian outgroups, we examined LDLR synteny in other sauropsids (alligators, turtles, lizard), frog and fish, in comparison to mammals (opossum, human). For analyses of predicted protein domains, we focused on avian species with complete genomic sequence and no gaps in LDLR and syntenic region. For zebra finch LDLR (Gene ID: 115491721), we used the mRNA/protein predictions (XM_030260053.1/XP_030115913.1) from a previous assembly (bTaeGut1_v1.p, (22)). Notably, the prediction in the current assembly (XM_032745657.1/XP_032601548.1 in bTaeGut2.pat.W.v2, (20)) contains a stop codon in exon 9 that prematurely truncates the predicted peptide, rendering the two 3' exons fully non-coding. A comparison with zebra finch transcriptomic data available in NCBI (SRA datasets) indicates the presence of an extra C (in a stretch of Cs) in XM_032745657.1 that is likely the result of an error in the assembly sequence; this mismatch compared to transcriptome data does not occur in XM_030260053.1. To confirm the absence of mammalian exons 2, 3 and 8 in songbirds, we performed BLAST searches of the relevant introns (1 and 5) in the zebra finch LDLR gene using avian (kakapo, golden eagle, chicken) sequences for these exons as queries, as well as translated these introns in all three frames. We found no significant hits or evidence of relevant orfs. We also examined the alignments of transcriptome data from a variety of tissues available for zebra finch at NCBI, and while there was variable but significant expression for all predicted 15 exons of zebra finch LDLR across tissues, we found no consistent evidence for additional exons within introns 1 and 5 of the zebra finch LDLR gene that might correspond to exons 2, 3 and 8 of LDLR in mammals. We confirmed the absence of these mammalian exons and corresponding protein domains in the LDLR from the high quality assemblies of the Swainson's thrush (*C. ustulatus*; Gene ID: 117009589 in bCatUst1.pri) and New Caledonian crow (*C. moneduloides*, Gene ID: 116437006 in bCorMon1.pri). Similarly, LDLR in the other two oscine species examined also lacked the same specific domains. In the case of the small tree finch (*C. parvulus*, Gene ID: 115917142 in STF_HiC) the gene prediction is incomplete, lacking the two 5'-most exons, even though there are no sequence gaps in this region. However, these exons are present, as assessed by BLAST searches (which revealed these exons are erroneously assigned to the upstream SMARCA4 gene), and yet the CRs 1-2 are missing in between them. In the case of the Tibetan ground tit (*P. humilis*, Gene ID: 106628808 in PseHum1.0), there are some local gaps in the assembly (Illumina-based) but no gaps between exons 1-2, and yet CRs 1-2 are missing. Thus, the genomic analysis conclusions are consistent across all five oscine species examined. For protein analysis, alignments of predicted peptides were generated using Clustal with default settings in JalView1.0 (https://osf.io/qdzg7/). Predicted structural domains in LDLR proteins were identified using

InterProScan. The human and finch sequences used for this comparison are available in supplementary Dataset 1. We note that several of the high quality assemblies Pacbio-based assemblies analyzed were made available through the Vertebrate Genome Project (VGP), as described in (20).

Animal and tissue preparation: Zebra finch (*Taeniopygia guttata*) blood and tissue samples were obtained from birds reared in our breeding colonies and used in brain studies. Embryos for LV injections and primary cell cultures were obtained from eggs laid in our breeding colonies. Chicken (Gallus gallus) eggs and blood samples were obtained from poultry sources (OR, USA and RN, Brazil). Animal protocols required for this work were approved by OHSU, Caltech and UFRN IACUC's internal review boards and are in accordance with NIH Guidelines.

Primary cultures: Zebra finch and chicken fibroblasts were dissociated from 3-day-old embryos and cultured in DMEM plus 10% FBS. These primary cultures were passed every 48 hr. Once established, fibroblasts were subjected to multiple rounds of infection by lentivirus expressing human LDLR (hLDLR, NCBI number Y114155.1) fused to GFP under the control of the ubiquitous promoter RSV. After 4-5 days, monolayers of wild type (wt) or hLDLR-expressing fibroblasts were incubated with VSV G LV particles containing red fluorescent protein with a nuclear localization signal (NLS-RFP) under RSV, and after a few days were examined for nuclear red fluorescence. VSV-G LV infectivity (infective units/µL) was measured by conducting a serial endpoint dilution assay. Data were analyzed by t-test (n=3 serial dilution plates per group, the experiment was repeated 3 times).

Embryo injections: Embryo injections were performed as previously described (11). Briefly, a small window was opened in the shell of freshly laid, fertilized eggs, and the embryo was visually located. Injections were performed with a hydraulic system using pulled glass pipettes with beveled tips to facilitate penetration. Embryos of both species (7 chicken and 8 finch embryos) were injected with 500 nL of the same viral preparation (LV-RSV-GFP at 10^8 IFU/µL). At the end of the injections, the egg window was sealed and the eggs placed back in an artificial incubator at 37°C and 50% humidity (Grumback, Germany). Embryos were collected 72 hr after injections and inspected for GFP expression under a fluorescence microscope.

Lipids assessment: Blood from normally reared adult male zebra finches was collected in the morning right after lights ON and before birds were allowed to feed (fasting; n=7), or after allowing a 30 min period of *ad lib* feeding (sated; n= 7). Blood was also collected from poultry roosters before or after they were allowed to feed in the morning (n= 3 each). For quantification of total serum cholesterol and triglyceride concentrations, fasting and sated serum samples were analyzed in duplicate using cholesterol liquid reagents (Pointe Scientific, Canton, MI) and triglycerides liquid reagents (Pointe Scientific, Canton, MI), respectively. For cholesterol, absorbance was measured at 490 nm, and for triglycerides, absorbance was measured at 540 nm using a microplate reader (SpectraMax iD3, Molecular Devices, San Jose, CA). Cholesterol and glycerol standards were used to determine serum concentrations (Pointe Scientific, Canton, MI). The serum lipoprotein distribution was determined by size exclusion chromatography [fast protein liquid chromatography (FPLC)] as previously described (3). Fractions obtained from FPLC were subjected to cholesterol and triglyceride analyses via the same colorimetric assays described above for total cholesterol and triglycerides. We note that for the FPLC

analysis, the serum samples from sated and fasting zebra finches were combined into 2 pooled samples (from 3 and 4 birds per group). The separate data from sated and fasting groups are shown in Figs. S3A-B, and the combined data for all finches (4 pooled samples) are shown in Fig. 2D (means +/- sem). All lipid-related analyses were performed at the Lipoprotein Analytical Core of OHSU's Knight Cardiovascular Institute.

Protein identification by Liquid Chromatography–Mass Spectrometry Pooled zebra finch serum was subjected to FPLC as described in *Lipids assessment* above, fractions of interest were combined, and 10 µg of protein was separated by 4-12% Bis-Tris Novex NuPage gel, with molecular weight assessed by SeeBlue Plus Pre-stained protein standard (Thermo Fisher). Coomassie staining was performed with the Power Blot system according to the manufacturer's protocol (Thermo Fisher). Bands of interest were excised and cut into 1 mm^3 cubes. In-gel tryptic digestion was carried out as previously described (4). In brief, gel pieces were rehydrated in 100 mM ammonium bicarbonate, and protein reduced with 10 mM dithiothreitol, followed by cysteine alkylation with 55 mM iodoacetamide, and digested overnight at 37°C with 1 µg trypsin (Thermo Fisher). Peptides were extracted in 5% formic acid followed by equal volume 100% acetonitrile and dried. Peptides were resuspended in 0.1% trifluoroacetic acid (TFA) and separated by a gradient on PepMap RSLC C18, 2 µm, 75 µm x 25 cm EasySpray column (Thermo Scientific). Data was collected on an Orbitrap Fusion Tribrid instrument configured with an EasySpray NanoSource (Thermo Scientific). RAW files were searched against a Uniprot database for *Taeniopygia guttata* (downloaded March 2017) with the SEQUEST HT search engine in Proteome Discoverer. Searches were configured with static modifications for carbamidomethyl (+57.021 Da) on cysteines, dynamic modifications for oxidation of methionine residues (+15.9949 Da), parent ion tolerance of 1.25 Da, fragment mass tolerance of 1.0005 Da, monoisotopic masses, and trypsin cleavage (max 2 missed cleavages). The large parent ion tolerance was used to increase the number of peptides being scored to improve discrimination of true versus false identifications. Searches used a reversed sequence decoy strategy to control peptide false discovery and identifications were validated by Percolator software (5). Only peptides with q scores $\leq$ 0.05 were accepted, and only 1 unique peptide was required for matching a protein entry for its identification. Uncharacterized entries were further characterized by sequence homology using BLASTP with a threshold of 0.001.

Online Methods References

22. Korlach J*, et al.* (2017) De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *Gigascience* 6(10):1-16.
23. Tavori H*, et al.* (2013) Serum proprotein convertase subtilisin/kexin type 9 and cell surface low-density lipoprotein receptor: evidence for a reciprocal regulation. *Circulation* 127(24):2403-2413.
24. Shevchenko A, Tomas H, Havlis J, Olsen JV, & Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1(6):2856-2860.
25. The M, MacCoss MJ, Noble WS, & Kall L (2016) Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom* 27(11):1719-1727.