Antiviral activity of a human placental protein of retroviral origin

John A. Frank[1], Manvendra Singh[1], Harrison B. Cullen[1], Raphael A. Kirou[1], Carolyn B. Coyne[2], Cédric Feschotte[1]

Affiliations

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

[2]Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, USA

Summary

Viruses circulating in wild and domestic animals pose a constant threat to human health[1]. Identifying human genetic factors that protect against zoonotic infections is a health priority. The RD-114 and Type-D retrovirus (RDR) interference group includes infectious viruses that circulate in domestic cats and various Old World monkeys (OWM), and utilize ASCT2 as a common target cell receptor[2]. While human ASCT2 can mediate RDR infection in cell culture, it is unknown whether humans and other hominoids encode factors that restrict RDR infection in nature[2,3]. Here we test the hypothesis that Suppressyn, a truncated envelope protein that binds ASCT2 and is derived from a human endogenous retrovirus[4,5], restricts RDR infection. Transcriptomics and regulatory genomics reveal that Suppressyn expression initiates in the preimplantation embryo. Loss and gain of function experiments in cell culture show Suppressyn expression is necessary and

sufficient to restrict RDR infection. Evolutionary analyses show Suppressyn was acquired in the genome of a common ancestor of hominoids and OWMs, but preserved by natural selection only in hominoids. Restriction assays using modern primate orthologs and reconstructed ancestral genes indicate that Suppressyn antiviral activity has been conserved in hominoids, but lost in most OWM. Thus in humans and other hominoids, Suppressyn acts as a restriction factor against retroviruses with zoonotic capacity. Transcriptomics data predict that other virus-derived proteins with potential antiviral activity lay hidden in the human genome.

Main

Viral zoonosis poses a constant threat to human health and has led to devastating epidemics such as those caused by Influenza[6], HIV[7], Ebola[8], and SARS coronaviruses[9,10]. Some zoonotic viruses have gained access to new host species by acquiring envelope (env) glycoproteins that mediate target-cell entry by binding to host cell surface receptors[6,11]. Notably, beta- and gamma-retroviruses have captured the so-called RDR env, which has enabled them to infect and transfer across diverse mammalian hosts[11,12]. For instance, the feline leukemia virus RD-114, an infectious endogenous retrovirus from the domestic cat, emerged from the Felis catus endogenous virus by acquiring the RDR *env* from the Baboon endogenous virus[11]. Because all RDR env bind to the highly conserved and broadly expressed amino acid transporter ASCT2 (also known as SLC1A5) to mediate cell entry, RDR env-mediated infection poses a zoonotic threat to humans[2,13,14]. Thus, it is critical to assess whether humans are equipped with mechanisms to protect against RDR infection.

Previous reports have shown that endogenous retroviral env are capable of restricting retroviral infection by a mechanism of receptor interference in multiple vertebrates, including chicken, mouse, sheep, and cat[15]. Some of these env-derived restriction factors acquired truncating mutations resulting in loss of their C-terminal membrane-anchoring transmembrane domain[15], but retention of receptor-binding activity. Suppressyn (*SUPYN)* is a protein that is derived from a human endogenous retroviral env, lacks a transmembrane domain, and is expressed throughout human placenta development[4,5]. Previous *in vitro* studies have shown that SUPYN, like Syncytin-1 (SYN1), binds ASCT2 and thereby modulates the fusogenic activity of SYN1 during placenta development[4,5]. Here we investigate whether *SUPYN* confers resistance to RDR infection.

SUPYN embryonic expression is driven by pluripotency and placentation regulatory factors.

To obtain a detailed view of *SUPYN* expression and regulation during human embryonic development, we analyzed publicly available scRNA-seq, ATAC-seq, DNAse-seq and ChIP-seq datasets generated from preimplantation embryos and human embryonic stem cells (hESC) **(Supplementary Table 1).** *SUPYN* mRNA appears after the onset of embryonic genome activation at the eight-cell stage and peaks in morula **(Fig 1a).** By blastula formation, *SUPYN* expression persists in the inner cell mass, epiblast, ESCs, and in the trophectoderm, which gives rise to the placenta **(Fig 1a, Extended Data 1a)**. Consistent with this expression pattern, the *SUPYN* locus is marked by open chromatin from 8-cell to blastocyst stages **(Extended Data 1b)**. In hESCs, the *SUPYN* promoter

region is marked by H3K4me1 and H3K27ac histone modifications characteristic of 'active' chromatin, and bound by core pluripotency (OCT4, NANOG, KLF4, SMAD1) and self-renewal (SRF, OTX2) transcription factors **(Fig 1b)**. Together, these data indicate *SUPYN* is robustly expressed throughout early embryonic development and likely under the control of pluripotency factors. By contrast, we found little evidence for *SYN1* expression in preimplantation embryos and hESCs **(Fig 1a, Extended Data 1a)**.

To examine *SUPYN* expression throughout placentation, we interrogated RNA-seq and ChIP-seq datasets generated from in vitro trophoblast (TB) differentiation models and placenta explants isolated at multiple stages of pregnancy **(Supplementary Table 1)**. During hESC to TB differentiation, we observed that pluripotency factors NANOG and OCT4 occupying the *SUPYN* promoter region are replaced by trophoblast-specific transcription factors TFAP2A and GATA3 **(Fig 1b).** *SUPYN* expression likely persists through the TB differentiation process because *SUPYN* transcripts and active chromatin marks (H3K27ac, H3K4me3, H3K9ac) are maintained across all analyzed TB cell lineages **(Fig 1b)**. By contrast, expression of other envelope-derived genes *SYN1*, *SYN2*, and *ERVV1*/*V2* is only detectable in differentiated trophoblasts **(Extended Data 1c).** We next mined scRNA-seq data generated from placenta at multiple developmental stages to examine the cell-type specific expression of *SUPYN* **(Supplementary Table 1)**. After classifying cell clusters based on known markers **(Fig 1c, d, Extended Data 2a, b, c)**, we found *SUPYN* and *ASCT2* expression specifically in the TB lineage **(Fig. 1e, f; Extended Data 2c, d)**. Consistent with previous reports[5], *SUPYN* expression was relatively high in cytotrophoblasts (CTB) and extra-villous trophoblasts (EVTB), but also

detectable in syncytiotrophoblasts (STB) **(Fig 1e, Extended Data 2c, d)**. *SUPYN* expression in EVTB was maintained throughout placental development **(Fig 1f)**. *SYN1* expression appears restricted to CTB and STB lineages **(Fig 1e, Extended Data 2c, d),** as previously reported[5,16-18]**.** To confirm these transcriptomic observations, we performed immunostaining of second (21w gestation) and third (31w gestation) trimester placenta with SUPYN antibody. These stains show SUPYN is widely expressed in STB, and perhaps CTB within the lumen of 2nd trimester placental villi **(Fig 1g, Extended Data 3).** Together these analyses indicate *SUPYN* is expressed throughout human fetal development and has a broader expression pattern than *SYN1.*

SUPYN confers resistance to RD114env-mediated infection

SUPYN expression during early embryonic and placental development, which coincides with that of ASCT2 **(Fig 1a),** suggests SUPYN may interact with ASCT2 throughout fetal development and confer RDR resistance to the developing embryo. To begin testing this hypothesis, we first examined whether human placenta-derived cell lines Jar and JEG3, and hESC H1 cells are resistant to RDR env-mediated infection. We generated HIV-GFP viral particles pseudotyped with either the feline RD114env (HIV-RD114env) or the glycoprotein G of vesicular stomatitis virus (HIV-VSVg, which uses the LDL receptor[19]) to monitor the level of infection in cell culture based on GFP expression **(Fig 2a, Extended Data 4)**[20]. These experiments revealed that Jar, JEG3, and H1 cells were susceptible to HIV-VSVg, as previously reported[21-25], but resistant to HIV-RD114env infection **(Fig 2b, c).** Concurrently infected 293T cells, which do not express *SUPYN* **(Extended Data Fig 1a)**, were similarly susceptible to infection by HIV-RD114env and HIV-VSVg **(Fig 2b, c)**.

116

117 To test whether SUPYN contributes to the HIV-RD114env resistance phenotype, we
118 repeated these infection experiments in Jar cells engineered to stably express short
119 hairpin RNAs depleting ~80% of *SUPYN* or *SYN1* mRNAs **(Extended Data Fig 5a).**
120 Depletion of *SUPYN* in Jar cells (shSUPYN) resulted in a significant increase in
121 susceptibility to HIV-RD114env infection **(Fig 2d)**, but did not affect infection by HIV-VSVg
122 **(Fig 2d**). Importantly, *SYN1* depletion from Jar cells did not increase susceptibility to HIV-
123 RD114env infection **(Fig 2d**). These results suggest that *SUPYN* expression contributes
124 to the RD114 resistance phenotype of Jar placental cells.

125

126 To account for possible off-target effects of *SUPYN*-targeting siRNAs, we transfected Jar-
127 shSUPYN cells with two siRNA-resistant, HA-tagged SUPYN rescue constructs and
128 examined their susceptibility to HIV-RD114env infection. Briefly, the siRNA-targeted
129 SUPYN signal peptide sequence was replaced with a luciferase (SUPYN-lucSP) or
130 modified signal peptide sequence (SUPYN-rescSP) that disrupts siRNA-binding but
131 retains codon-identity. Transfection with either SUPYN-rescSP or SUPYN-lucSP restored
132 a significant level of resistance to HIV-RD114env infection **(Fig 2e).** Western Blot analysis
133 of transfected cell lysates showed SUPYN-rescSP was more abundantly expressed than
134 SUPYN-lucSP **(Fig 2f)**, which may account for the stronger HIV-RD114env resistance
135 conferred by SUPYN-rescSP **(Fig 2e).** These results corroborate the notion that SUPYN
136 restricts RD114env-mediated infection in Jar cells.

137

To test if SUPYN expression alone is sufficient to confer protection against RD114env-mediated infection, we transfected 293T cells with a SUPYN overexpression construct and subsequently infected with HIV-RD114env and HIV-VSVg respectively. As a positive control, we also transfected 293T cells with a RD114env overexpression construct, which is predicted to confer resistance to HIV-RD114env, but not to HIV-VSVg. Expression of either SUPYN or RD114env resulted in ~80% reduction in the level of HIV-RD114env infection **(Fig 2g, Extended Data Fig 6a)**, but had no significant effect on HIV-VSVg infectivity **(Extended Data Fig 6b)**. Taken together, our knockdown and overexpression experiments indicate SUPYN expression is both necessary and sufficient to confer resistance to RD114env-mediated infection.

SUPYN restricts RDR infection likely through receptor interference

Our RD114env-specific resistance phenotype **(Extended Data Fig 6a, b)** strongly suggests SUPYN functions by receptor interference. If so, this protective effect should extend to infection mediated by other RDR env[2,26,27] since they all use ASCT2 as receptor. To test this prediction, we generated HIV-GFP reporter virions pseudotyped with Squirrel Monkey Retrovirus (SMRV) env (HIV-SMRVenv)[2] **(Fig 2a)** and infected 293T cells previously transfected with SUPYN, SMRVenv (as a positive control) or an empty vector. Cells expressing SUPYN or SMRVenv showed an ~80% reduction of HIV-SMRVenv infected cells **(Fig 2h)**. Thus, SUPYN can restrict infection mediated by multiple RDRenv.

Another prediction of RDR restriction via receptor interference is that it should be a property of env proteins recognizing ASCT2, such as SUPYN, but not those binding other

cellular receptors. Consistent with this prediction, expressing HA-tagged env from amphotrophic murine leukemia virus (aMLV) or human endogenous retrovirus H, neither of which are expected to interact with ASCT2[28-30], had no effect on HIV-RD114env nor HIV-VSVg infection in 293T cells, while HA-tagged SUPYN strongly restricted HIV-RD114env **(Extended Data Fig 6a, b)**. Importantly, all tested env proteins were expressed at comparable levels **(Extended Data Fig 6c).** Furthermore, we observed that SUPYN overexpression did not significantly impact ASCT2 expression levels in 293T cells **(Extended Data Fig 6c).** This result suggests that if SUPYN acts by receptor interference, its interaction with ASCT2 does not result in receptor degradation, which is consistent with some instances of receptor interference[31-33]. In agreement with previous observations[5], we noted that *SUPYN* knockdown in Jar cells resulted in the specific loss of a putative non-glycosylated isoform of ASCT2 **(Extended Data 5b)**. We speculate that the presence of SUPYN-dependent non-glycosylated ASCT2 may be the result of SUPYN sterically interfering with the glycosylation machinery within the secretory pathway. It has not been reported whether ASCT2 glycosylation impacts RDR env-mediated infection in human cells, but it is known that receptor glycosylation may interfere with RDR infection in mouse and hamster cells[34,35]. Collectively, these observations converge on the model that SUPYN restricts against RDR infection through receptor interference **(Fig 2i)**.

***SUPYN* emerged in a catarrhine ancestor and evolved under functional constraint in hominoids**

Little is known about the evolutionary origin of *SUPYN*. The gene was originally identified as derived from a copy of the HERV-Fb family of endogenous retroviruses (also known

as HERVH48[36]) located on human chromosome 21q22.3 with an ortholog in chimpanzee[4]. Using comparative genomics (see Methods), we found that this HERVH48 element is shared at orthologous position across the genomes of all available hominoids (i.e. apes) and most Old World monkeys (OWM), but precisely lacking in New World monkeys and prosimians **(Fig 3a, Extended Data 7, 8).** Thus, the provirus copy that gave rise to *SUPYN* inserted in the common ancestor of catarrhine primates ~20-38 million years ago[37] **(Fig 3a)**.

All primates with HERVH48 orthologs also share a nonsense mutation which would have truncated the ancestral env protein at site 185 in the common ancestor of catarrhine primates **(Fig 3a, Extended Data 8)**. Hominoids share an additional nonsense mutation further truncating the protein to the 160-aa SUPYN-encoding ORF annotated in the human reference genome **(Fig 3a, Extended Data 8).** The *SUPYN* ORF is almost perfectly conserved in length across hominoids, but not in OWM where some species display further frameshifting and truncating mutations **(Extended Data 9)**, suggesting *SUPYN* may have evolved under different evolutionary regimes in hominoids and OWMs. To test this idea, we analyzed the ratio (ω) of nonsynonymous (dN) to synonymous (dS) substitution rates using codeml[38], which provides a measure of selective constraint acting on codons. Log-likelihood ratio tests comparing models of neutral evolution with selection indicate *SUPYN* evolved under purifying selection in hominoids ($\omega = 0.38$; p = 1.47E-02), but did not depart from neutral evolution in OWMs ($\omega = 1.44$; p = 0.29) **(Fig 3a).** For comparison, we performed the same type of analysis for *SYN1* and *SYN2*, two primate-specific *env*-derived genes thought to be involved in placentation[17,39,40]. Consistent with

previous reports[41,42], we found that both *SYN1* ($\omega = 0.64$; p = 1.80E-02) and *SYN2* (ω = 0.29; p = 3.22E-08) evolved under purifying selection during hominoid evolution **(Fig. 3a)**. In OWMs, *SYN2* also evolved under purifying selection (ω = 0.22, p = 2.78E-08), while *SYN1* was lost through an ancestral deletion[14] **(Fig. 3a).** These results suggest that the level of functional constraint acting on *SUPYN* during hominoid evolution is comparable to that seen on other *env*-derived genes with placental function.

SUPYN antiviral activity is conserved across hominoid primates

To assess whether primate *SUPYN* orthologs have antiviral activity, we generated and transfected 293T cells with HA-tagged overexpression constructs for the orthologous SUPYN sequences of chimpanzee, siamang, African green monkey, pigtailed macaque, crab-eating macaque, rhesus macaque, and olive baboon and challenged these cells with HIV-RD114env virions. Both chimpanzee and siamang SUPYN proteins displayed antiviral activity with potency comparable to or greater than human SUPYN, respectively **(Fig 3b).** By contrast, only one (African green monkey) of the five OWM orthologous SUPYN proteins exhibited a modest but significant level of antiviral activity **(Fig 3b, c).** The lack of restriction activity for some of the OWM proteins may be attributed to their relatively low expression level in these human cells **(Fig 3b)** and/or their inability to bind the human ASCT2 receptor due to *SUPYN* sequence divergence **(Extended Data 8, 9)**. To gain further insight into the evolutionary origins of SUPYN antiviral activity, we expressed SUPYN sequences predicted for the common ancestor of hominoid and OWM **(Extended Data 9)** (see Methods) and assayed their antiviral activity in 293T cells. Both ancestral proteins were expressed at levels comparable to human and green monkey

SUPYN and exhibited strong restriction activity **(Fig 3c).** These data indicate that SUPYN antiviral activity against RDR env-mediated infection is an ancestral trait preserved over ~20 million years of hominoid evolution, but apparently lost in some OWM lineages.

Truncated HERV env-coding sequences are expressed in diverse cell types

The antiviral activity of SUPYN against RDR env-mediated infection raises the possibility that the human genome may harbor other *env*-derived protein-coding genes with antiviral function. To assess the potential pool of such sequences, we scanned the human genome for *env*-derived open reading frames (*envORF)* that minimally encode at least 70 aa predicted to include the receptor-binding surface domain (see Methods). This search identified a total of 1,507 unique *envORFs,* including ~20 env-derived sequences currently annotated as human genes such as *SUPYN* and *SYN1*[4,43,44]. We then mined transcriptome datasets generated from human preimplantation embryos and various tissues **(Supplementary Table 1)**, and observed that ~44% (668/1507) of *envORF*s showed evidence of RNA expression in at least one of the cell types surveyed **(Fig 4).** These analyses revealed three general insights about expressed non-annotated *envORFs*: (1) like known *env*-derived genes, *envORFs* exhibit tissue-specific expression patterns **(Fig 4; Extended Data 11, 12)**; (2) the majority of *envORFs* are expressed during human fetal development or in stem and progenitor cells **(Fig 4; Extended Data 11)**; (3) *envORFs* are rarely expressed in differentiated tissues under normal conditions, with the exception of brain **(Fig 4; Extended Data 11, 12)**. These analyses suggest that the human genome harbors a vast reservoir of *env*-derived sequences with coding and

receptor-binding potential, many of which are transcribed in a tissue-specific fashion and may encode receptor-binding antiviral peptides.

Discussion

Our expression and selection analyses firmly establish that *SUPYN* is a bona fide gene encoding a truncated envelope of retroviral origin that is expressed in the human preimplantation embryo and throughout placental development. Virologic assays in human cell culture show SUPYN is necessary and sufficient to confer resistance to RDR env-mediated infection, likely by competing for the receptor (ASCT2) utilized by this diverse set of retroviruses **(**see model**, Fig 2i)**. The expression profile of *SUPYN* and the RD114 resistance phenotype of human ESCs and placental cells suggest *SUPYN* provides a layer of protection against RDR infection to the developing embryo, including the nascent germline. The observation that infectious and replication-competent RDRs are known to currently circulate in several mammals, both as exogenous or endogenous viruses, but not in hominoids[2] lends further support to a model in which SUPYN has contributed to hominoid resistance to RDR infection and possibly endogenization.

Like *SYN1*, *SUPYN* emerged in the common ancestor of catarrhine primates and was preserved by natural selection in hominoids. The parallel evolutionary path of *SYN1* and *SUPYN* and their pattern of expression in the placenta are compatible with a model in which *SUPYN* acts as a negative modulator of *SYN1* fusogenic activity during STB development[4,5]. The developmental and antiviral functions of SUPYN are not mutually

exclusive, but may even be interlocked. Syncytins, including SYN1, are fully functional envelopes that can be incorporated into heterologous retroviral particles and exosomes originating from the placenta[50-55,45]. Because ASCT2 is broadly expressed, SYN1-pseudotyped particles produced in the developing placenta have the potential to infiltrate a wide range of surrounding cell types. Thus, the physiological benefits afforded by Syncytins in promoting cell-cell fusion during STB development may have come with the cost of exposing the developing embryo to a wide variety of endogenous and exogenous invasive genetic elements, including but not limited to RDRs, that could be serendipitously enveloped by SYN1 throughout pregnancy. It is tempting to speculate that *SUPYN* has been maintained by natural selection to shield the developing embryo from the constant threat and adverse effects of SYN1-mediated infections. The conserved antiviral activity of ancestral hominoid and OWM *SUPYN* suggest resistance against RDR infection may have precipitated the initial retention of *SUPYN* in a catarrhine ancestor, and subsequently facilitated the domestication of *SYN1* in hominoids.

More broadly, this study serves as a proof of principle that truncated envelope peptides expressed from relics of ancient retroviruses integrated in the human genome can exert and retain antiviral activities for millions of years. We identified hundreds of candidate env-derived genes in the human genome that may encode peptides with receptor-binding activity and antiviral function. Furthermore, Gag (capsid)-derived proteins encoded by endogenous retroviruses are also capable of retroviral restriction[46]. Thus, it is possible that our genome holds a vast reservoir of retrovirus-derived proteins with protective activity against various zoonotic agents, including non-retroviral pathogens.

Main References

1. Warren, C. J. & Sawyer, S. L. How host genetics dictates successful viral zoonosis. *PLoS Biol.* **17,** e3000217 (2019).
2. Sinha, A. & Johnson, W. E. ScienceDirect Retroviruses of the RDR superinfection interference group: ancient origins and broad host distribution of a promiscuous Env gene. *Current Opinion in Virology* **25,** 105–112 (2017).
3. Montiel, N. A. An updated review of simian betaretrovirus (SRV) in macaque hosts. *J. Med. Primatol.* **39,** 303–314 (2010).
4. Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y. & Schust, D. A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci Rep* **3,** 1462–8 (2013).
5. Sugimoto, J. *et al.* Suppressyn localization and dynamic expression patterns in primary human tissues support a physiologic role in human placentation. *Sci Rep* **9,** 19502–12 (2019).
6. Mostafa, A., Abdelwhab, E. M., Mettenleiter, T. C. & Pleschka, S. Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses* **10,** 497 (2018).
7. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* **1,** a006841–a006841 (2011).
8. Baseler, L., Chertow, D. S., Johnson, K. M., Feldmann, H. & Morens, D. M. The Pathogenesis of Ebola Virus Disease. *Annu Rev Pathol* **12,** 387–418 (2017).
9. Song, Z. *et al.* From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses* **11,** 59 (2019).

10. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **100,** 1–3 (2020).

11. Henzy, J. E. & Johnson, W. E. Pushing the endogenous envelope. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368,** 20120506–20120506 (2013).

12. Diehl, W. E., Patel, N., Halm, K. & Johnson, W. E. Tracking interspecies transmission and long-term evolution of an ancient retrovirus using the genomes of modern mammals. *Elife* **5,** e12704 (2016).

13. Green, B. J., Lee, C. S. & Rasko, J. E. J. Biodistribution of the RD114/mammalian type D retrovirus receptor, RDR. *J Gene Med* **6,** 249–259 (2004).

14. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47,** D853–D858 (2019).

15. Johnson, W. E. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* **72,** 5955 (2019).

16. Holder, B. S., Tower, C. L., Abrahams, V. M. & Aplin, J. D. Syncytin 1 in the human placenta. *Placenta* **33,** 460–466 (2012).

17. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–789 (2000).

18. Frendo, J. L. *et al.* Direct Involvement of HERV-W Env Glycoprotein in Human Trophoblast Cell Fusion and Differentiation. *Molecular and Cellular Biology* **23,** 3566–3574 (2003).

19. Finkelshtein, D., Werman, A., Novick, D., Barak, S. & Rubinstein, M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 7306–7311 (2013).

20. Sandrin, V., Muriaux, D., Darlix, J. L. & Cosset, F. L. Intracellular Trafficking of Gag and Env Proteins and Their Interactions Modulate Pseudotyping of Retroviruses. *J. Virol.* **78,** 7153–7164 (2004).

21. Dottori, M., Tay, C. & Hughes, S. M. Neural development in human embryonic stem cells-applications of lentiviral vectors. *J. Cell. Biochem.* **112,** 1955–1962 (2011).

22. Gropp, M. *et al.* Stable genetic modification of human embryonic stem cells by lentiviral vectors. *Mol. Ther.* **7,** 281–287 (2003).

23. Sakata, M. *et al.* Analysis of VSV pseudotype virus infection mediated by rubella virus envelope proteins. *Sci Rep* **7,** 11607 (2017).

24. Delorme-Axford, E. *et al.* Human placental trophoblasts confer viral resistance to recipient cells. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 12048–12053 (2013).

25. Vidricaire, G., Tardif, M. R. & Tremblay, M. J. The low viral production in trophoblastic cells is due to a high endocytic internalization of the human immunodeficiency virus type 1 and can be overcome by the pro-inflammatory cytokines tumor necrosis factor-alpha and interleukin-1. *J. Biol. Chem.* **278,** 15832–15841 (2003).

26. Johnson, W. E. Endogenous Retroviruses in the Genomics Era. *Annu Rev Virol* **2,** 135–159 (2015).

27. Sommerfelt, M. A. & Weiss, R. A. Receptor interference groups of 20 retroviruses plating on human cells. *Virology* **176,** 58–69 (1990).

28. de Parseval, N., Casella, J.-F., Gressin, L. & Heidmann, T. Characterization of the Three HERV-H Proviruses with an Open Envelope Reading Frame

Encompassing the Immunosuppressive Domain and Evolutionary History in Primates. *Virology* **279,** 558–569 (2001).

29. van Zeijl, M. *et al.* A human amphotropic retrovirus receptor is a second member of the gibbon ape leukemia virus receptor family. *Proc Natl Acad Sci USA* **91,** 1168–1172 (1994).

30. Miller, D. G., Edwards, R. H. & Miller, A. D. Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukemia virus. *Proc Natl Acad Sci USA* **91,** 78–82 (1994).

31. Liu, M. & Eiden, M. V. The receptors for gibbon ape leukemia virus and amphotropic murine leukemia virus are not downregulated in productively infected cells. *Retrovirology* **8,** 53–14 (2011).

32. Jobbagy, Z., Garfield, S., Baptiste, L., Eiden, M. V. & Anderson, W. B. Subcellular redistribution of Pit-2 P(i) transporter/amphotropic leukemia virus (A-MuLV) receptor in A-MuLV-infected NIH 3T3 fibroblasts: involvement in superinfection interference. *J. Virol.* **74,** 2847–2854 (2000).

33. Kim, J. W. & Cunningham, J. M. N-linked glycosylation of the receptor for murine ecotropic retroviruses is altered in virus-infected cells. *J. Biol. Chem.* **268,** 16316–16320 (1993).

34. Marin, M., Tailor, C. S., Nouri, A. & Kabat, D. Sodium-dependent neutral amino acid transporter type 1 is an auxiliary receptor for baboon endogenous retrovirus. *J. Virol.* **74,** 8085–8093 (2000).

35. Marin, M., Lavillette, D., Kelly, S. M. & Kabat, D. N-linked glycosylation and sequence changes in a critical negative control region of the ASCT1 and ASCT2

neutral amino acid transporters determine their retroviral receptor functions. *J. Virol.* **77,** 2936–2945 (2003).

36. Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44,** D81–9 (2016).

37. Perelman, P. *et al.* A molecular phylogeny of living primates. *PLoS Genet* **7,** e1001342 (2011).

38. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24,** 1586–1591 (2007).

39. Malassiné, A. *et al.* Expression of the fusogenic HERV-FRD Env glycoprotein (syncytin 2) in human placenta is restricted to villous cytotrophoblastic cells. *Placenta* **28,** 185–191 (2007).

40. Mallet, F. *et al.* The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proc Natl Acad Sci USA* **101,** 1731–1736 (2004).

41. Bonnaud, B. *et al.* Evidence of selection on the domesticated ERVWE1 env retroviral element involved in placentation. *Molecular Biology and Evolution* **21,** 1895–1901 (2004).

42. de Parseval, N. *et al.* Comprehensive search for intra- and inter-specific sequence polymorphisms among coding envelope genes of retroviral origin found in the human genome: genes and pseudogenes. *BMC Genomics* **6,** 117–11 (2005).

43. Heidmann, O. *et al.* HEMO, an ancestral endogenous retroviral envelope protein shed in the blood of pregnant women and expressed in pluripotent stem cells and tumors. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E6642–E6651 (2017).

44. Heidmann, A. D. C. L. T., Lavialle, C. & Heidmann, T. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta* **33,** 663–671 (2012).

45. Tang, Y. *et al.* Endogenous Retroviral Envelope Syncytin Induces HIV-1 Spreading and Establishes HIV Reservoirs in Placenta. *Cell Rep* **30,** 4528–4539.e4 (2020).

46. Frank, J. A. & Feschotte, C. Co-option of endogenous viral sequences for host cell function. *Current Opinion in Virology* **25,** 81–89 (2017).

Figures

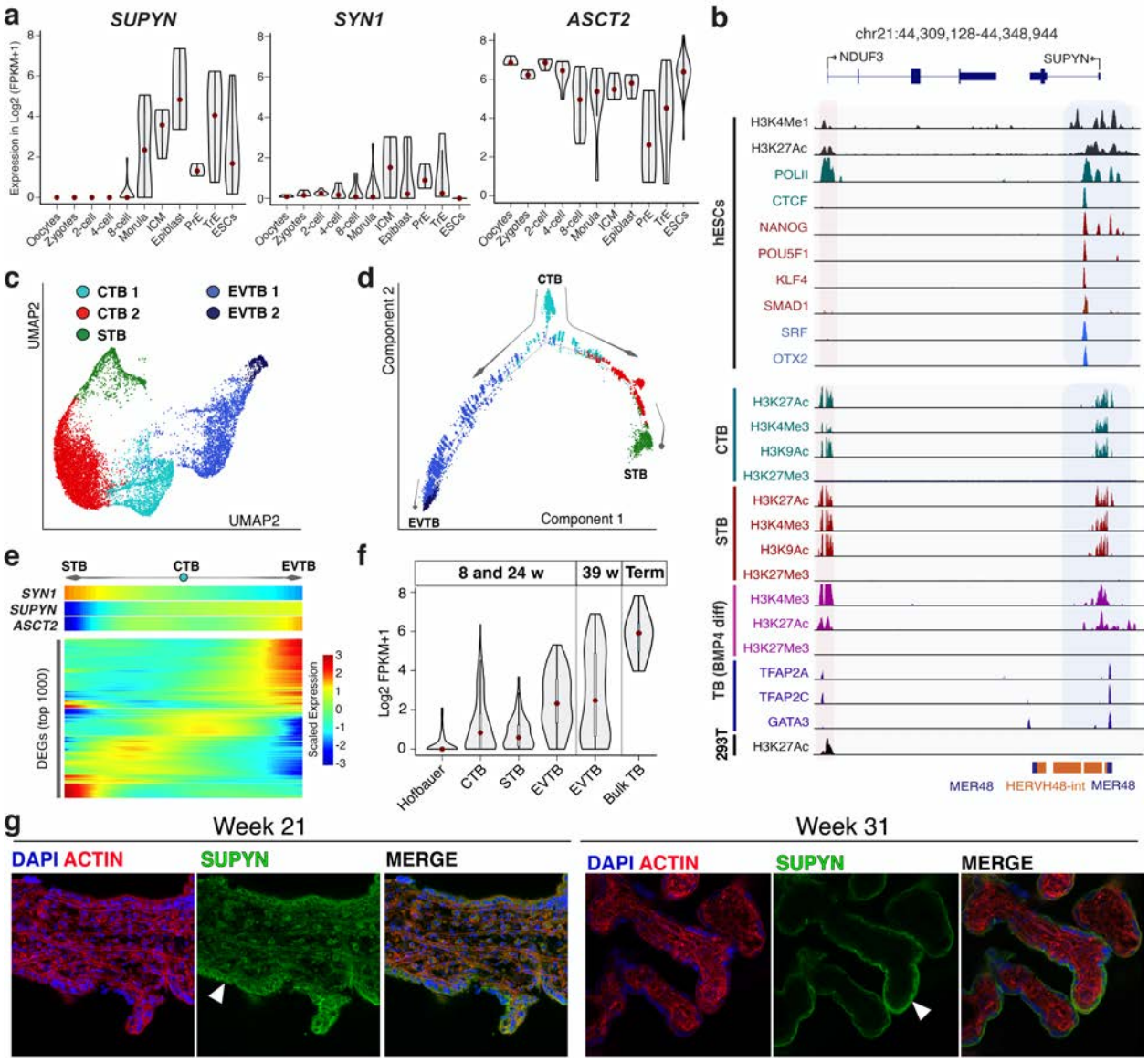


Figure 1: Pluripotency and placentation regulatory factors drive SUPYN expression during fetal development.

(a) Violin plots summarizing SUPYN, SYN1 and ASCT2 expression in human preimplantation embryo and ESC single-cell RNA-seq data. **(b)** Genome browser view of the regulatory elements around the SUPYN locus in hESCs, cyto- (CTB), syncytio- (STB), BMP4 differentiated trophoblasts (TB), and 293T cells. ChIP-seq profiles are shown for indicated transcription factors and histone

444 modifications with shaded area highlighting regions of active chromatin. **(c)** UMAP visualization
445 of TB cell clusters, shown in **Extended Data 2**. **(d)** Monocle2 pseudotime analysis of cell clusters
446 in **c** illustrates the developmental trajectory of CTBs that give rise to STB and EVTB respectively.
447 Color codes in **c** and **d** denote cell identity. **(e)** Heatmap represents the top 1000 differentially
448 expressed genes (row) of single cells (column) sorted according to pseudotime analyzed in **c** and
449 **d**. Cells are ordered according to the pseudotime progression of CTB (middle) to STB (left) and
450 EVTB (right). SYN1, SUPYN, and ASCT2 were fetched from the heatmap below. **(g)** Confocal
451 microscopy of placental villi explants stained for SUPYN (green), Actin (red) and DAPI. STBs are
452 marked by arrowheads.
453
454

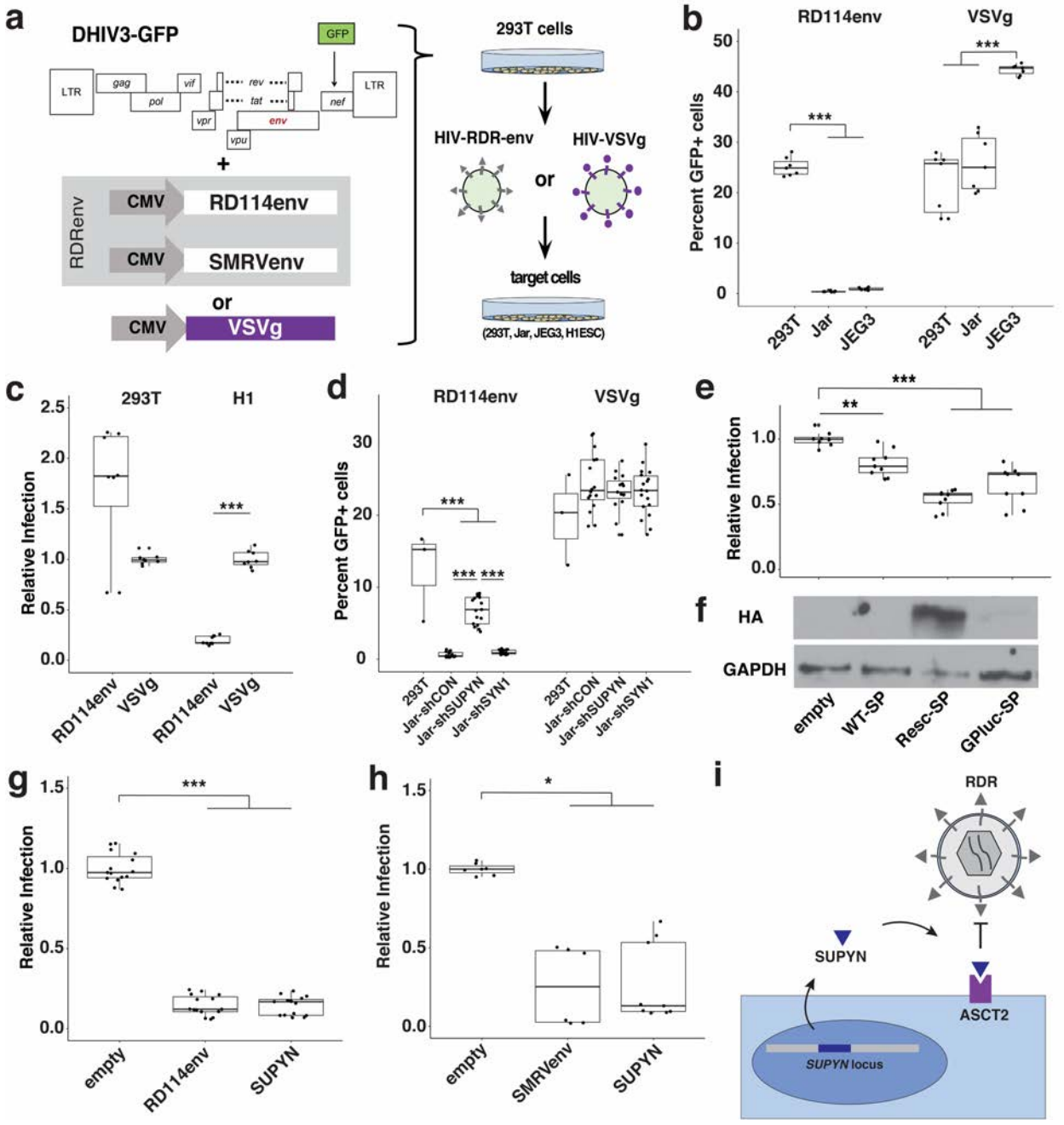
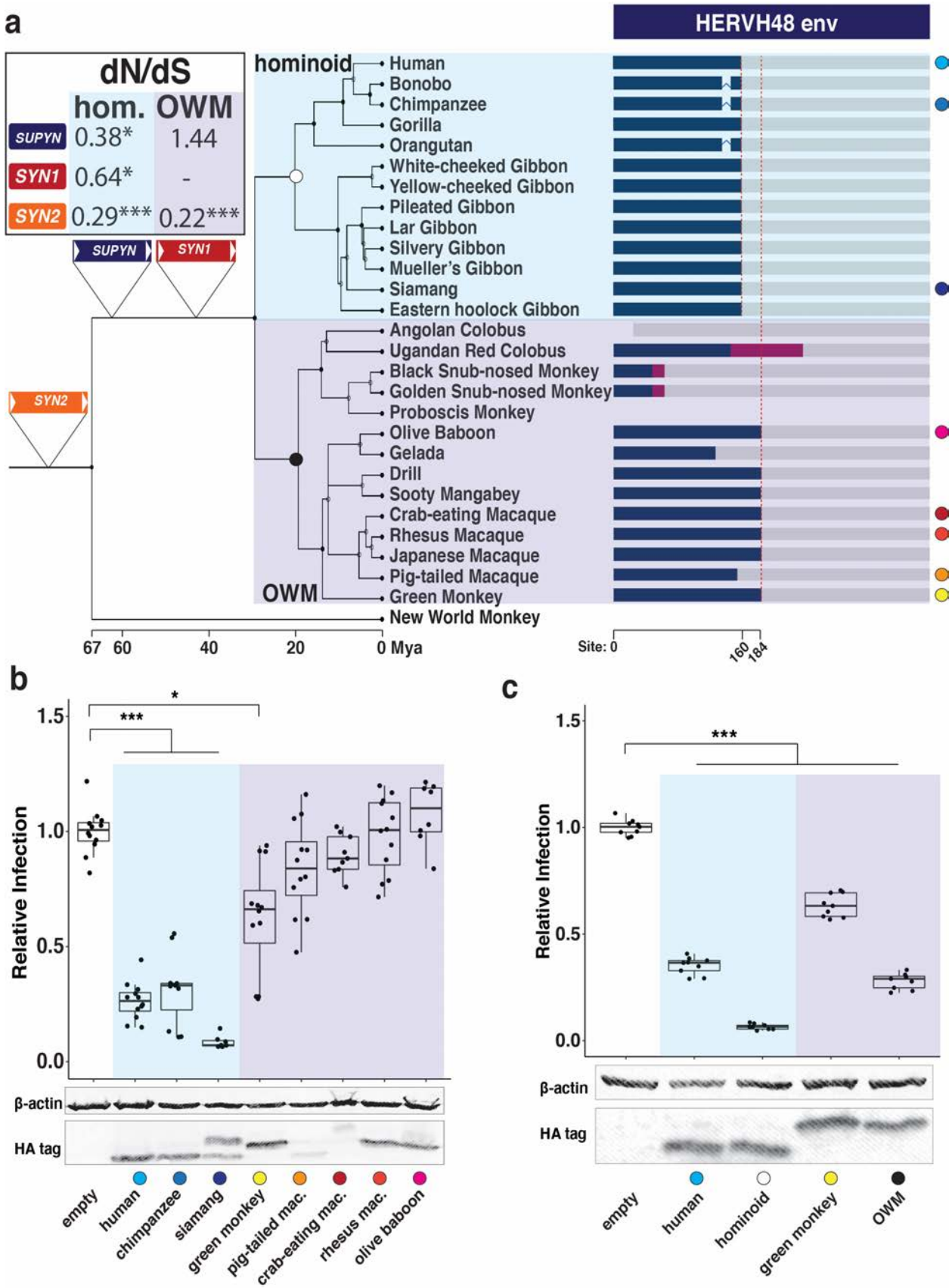


Figure 2: SUPYN confers resistance to RDR env-mediated infection

(a) Virus production and infection assay approach (see Methods). **(b, d)** Proportion of GFP+ 293T, JEG3, Jar, and shRNA-transduced Jar cells infected with HIV-RD114env or HIV-VSVg. **(c)** Relative infection rate of 293T and H1-ESCs normalized to mean proportion of HIV-VSVg-infected cells. **(e, g, h)** Relative infection rates of GFP+ 293T cells transfected with **(e)** wild-type (WT-SP), rescue

(Resc-SP), luciferase signal peptide (GPluc-SP), unmodified **(e, g)** SUPYN, **(g)** RD114env, or **(h)** SMRVenv overexpression constructs. Relative infection was determined by normalizing indicated constructs to empty vector ($n \geq 3$ with ≥ 2 technical replicates; ***adj. $p < 0.001$; **adj. $p < 0.01$; *adj. $p < 0.05$; ANOVA with Tukey HSD). **(f)** Western Blot analysis (αHA, αGAPDH) of 293T cell lysates transfected with indicated constructs. **(i)** Model of SUPYN-dependent RDR infection restriction.



468

469 **Figure 3: SUPYN emerged in a Catarrhine ancestor and has conserved antiviral activity in**
470 **Hominoids.**

471 **(a)** Consensus primate phylogeny with cartoon representation of *SUPYN* ORFs (blue box).
472 Magenta boxes represent frame-shifts in *SUPYN* ORFs. Red dashed lines denote conserved
473 premature stop codon positions. Grey bars represent degraded HERVH48env sequence. Labeled
474 triangles denote ancestral lineage where HERVH48env was acquired. Colored circles indicate
475 species used in **b** and **c.** *SUPYN*, *SYN1* and *SYN2* dN/dS values are shown in box (***$p < 0.001$; *p
476 < 0.05; LRT). **(b, c)** Relative infection rates and Western Blot of 293T cells transfected with primate
477 (**b**) or ancestral (**c**) SUPYN-HA constructs and infected with HIV-RD114env are shown. Relative
478 infection rates were determined by normalizing GFP+ counts to empty vector. ($n \geq 3$ with ≥ 2
479 technical replicates ***adj. $p < 0.001$; *adj. $p < 0.05$; ANOVA with Tukey HSD)

480

481

482

483

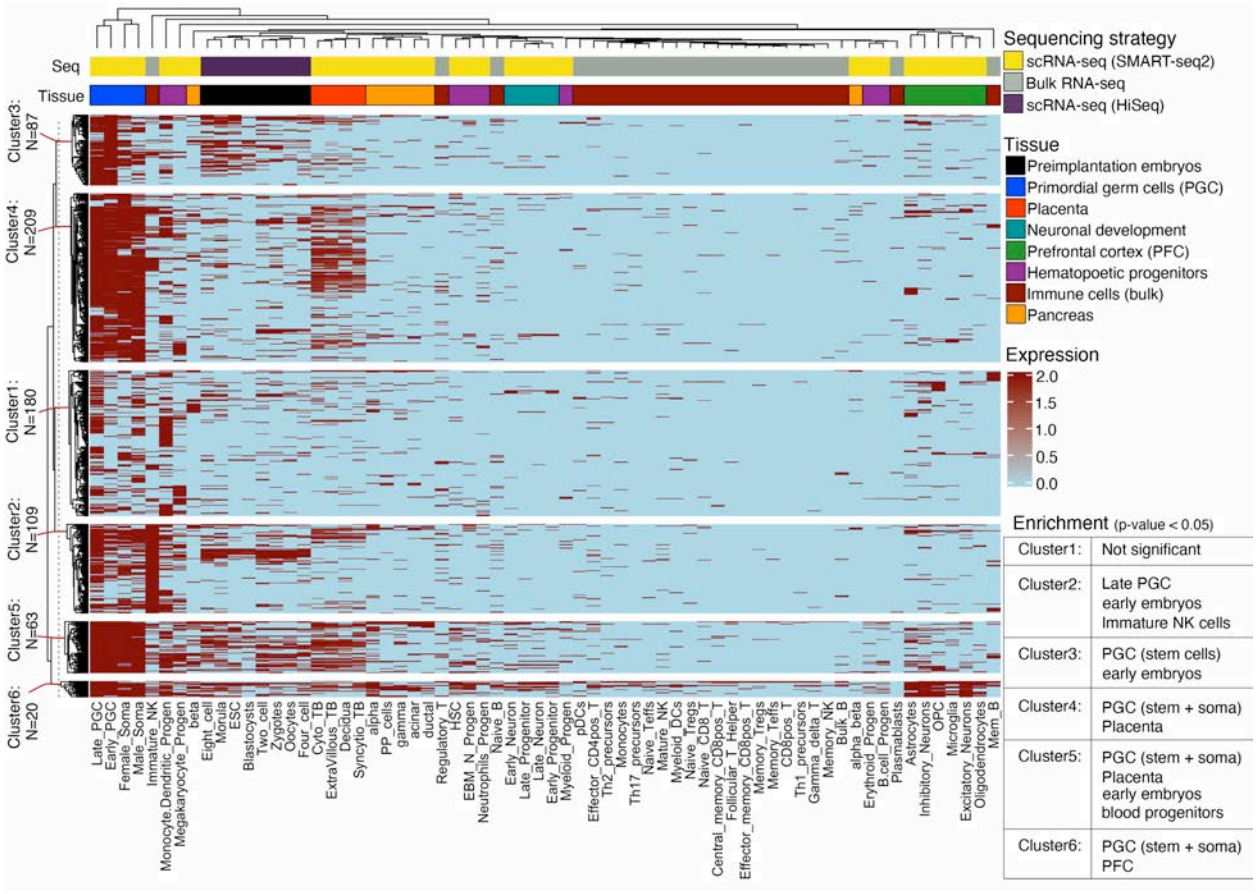


Figure 4: Expression profile of env-derived transcripts over a subser of human cell types.

Heatmap shows expression of 668 *envORF* loci in 66 distinct cell-types from 8 independent datasets shown in Extended Data 11. Rows represent individual *envORF* loci expressed (log2 CPM > 1) in at least one cell type (columns). Upper bars, located above the heatmap, denote the sequencing strategy (top) and tissue source (bottom) with the same color scheme shown in Extended Data 11. Rows are clustered into six distinct groups based on their expression. Significant envORF enrichment in cell-types was calculated using a hypergeometric test.

Methods

RNA-seq analyses

We mined published single cell transcriptome datasets (scRNA-seq) of human pre-implantation embryos isolated at developmental stages ranging from oocyte to blastocyst[47] (GSE36552) and from human placenta[48,49] (GSE89497, GSE87726). Reads were mapped to the human genome (hg19) with STAR[50] using the following settings *--alignIntronMin 20 --alignIntronMax 1000000 --chimSegmentMin 15 --chimJunctionOverhangMin 15 --outFilterMultimapNmax 20*. Only uniquely mapped reads were considered for expression calculations. Gene level counts were obtained using *featureCounts*[51] run with RefSeq annotations. Gene expression levels were calculated at Transcript Per Million (TPM) from counts mapped over the entire gene (defined as any transcript located between Transcription Start Site (TSS) and Transcription End Site (TES)). Only genes and cells that met the following criteria were included in this analysis: (1) each cell must express at least 5,000 genes; (2) each gene must be expressed in at least 1% of cells; (3) each gene must be expressed with log2 TPM >1. We clustered cells meeting these criteria using the default parameters of the Seurat[52] package (v3.1.1) implemented in R (v3.6.0). Seurat applies the most variable genes to get top principal components that are used to discriminate cell clusters in tSNE or UMAP plots. We set Seurat to use 10 principal components in this cluster analysis. For the placental scRNAseq data **(Fig1, Extended Data 2)**, the 2000 most differentially expressed genes were used to define cell clusters. Major clusters corresponding to CTB, STB, EVTB, macrophages, and stromal cells were identified based on the expression of known marker

518 genes. Monocle2[53] was used to perform single-cell trajectory analysis and cell ordering
519 along an artificial temporal continuum. The top 500 differentially expressed genes were
520 used to distinguish between CTB, STB and EVTB cell populations. The transcriptome
521 from each single cell represents a pseudo-time point along an artificial time vector that
522 denotes the progression of CTB to STB or EVTB respectively.
523 Data generated on the 10X Genomics scRNA-seq platforms were processed in the
524 following way. The processed data matrix from Vento-Tormo et al.[54] was retrieved from
525 the E-MTAB-6701 entry. The normalized counts and cell-type annotations were used as
526 provided by the original publications. Seurat was used for filtering, normalization and cell-
527 type identification. The following data processing steps were performed: (1) Cells were
528 filtered based on the criteria that individual cells must have between 1,000 and 5,000
529 expressed genes with a count ≥1; (2) cells with more than 5% of counts mapping to
530 mitochondrial genes were filtered out; (3) data was normalized by dividing uniquely
531 mapping read counts (defined by Seurat as unique molecular identified (UMI)) for each
532 gene by the total number of counts in each cell and multiplying by 10,000. These
533 normalized values were then natural-log transformed. Cell types were defined by using
534 the top 2000 variable features expressed across all samples. Clustering was performed
535 using the “FindClusters” function with largely default parameters; except resolution was
536 set to 0.1 and the first 20 PCA dimensions were used in the construction of the shared-
537 nearest neighbor (SNN) graph and the generation of 2-dimensional embeddings for data
538 visualization using UMAP. Cell types were assigned based on the annotations provided
539 by the original publication.

Bulk RNAseq datasets generated from placenta[55], 293T[56,57] and human immune cells[58], were processed as described above. Briefly, reads were mapped with STAR and uniquely mapped reads were counted with featureCounts.

ChIP-seq, DNAse-seq and ATAC-seq data analysis

Various ChIP-seq datasets representing histone modifications and transcription factors in human embryonic stem cells and their differentiation were retrieved from[59,60] (GSE61475, GSE99631). We obtained the H3K27Ac[61] (GSE127288) for CTB to STB primary cultures, H3K4Me1 for trophoblasts[62] (GSE118289), H3K4Me3, H3K27Me3 for differentiated trophoblasts[63] (GSE105258), and GATA2/3, TFAP2A/C[63] (GSE105081) ChIP-seq datasets in raw fastq format. DNAse-seq and ATAC-seq datasets were retrieved from Gao et al.[64] (GSA:CRA000297) and Wu et al.[65] (GSE101571) respectively.

Reads from the above described datasets were aligned to the hg19 human reference genome using Bowtie2[66] run in *--very-sensitive-local* mode. All reads with MAPQ < 10 and PCR duplicates were removed using Picard and *samtools*[67]. All of the peaks were called by MACS2[68] (https://github.com/macs3-project/MACS) with the parameters in narrow mode for TFs and broad mode for histone modifications keeping FDR < 1%. ENCODE-defined blacklisted regions[69] were excluded from called peaks. We then intersected these peak sets with repeat elements from hg19 repeat-masked coordinates using bedtools *intersectBed*[70] with a 50% overlap. To visualize over Refseq genes (hg19) using IGV[71], the raw signals of ChIP-seq were obtained from MACS2, using the parameters: *-g hs -q 0.01 -B*. The conservation track was visualized through UCSC

562 genome browser[14] under net/chain alignment of given non-human primates (NHPs) and
563 merged beneath the IGV tracks.
564
565 **Cell culture**
566 293T cells were cultured in DMEM (GIBCO, 11995065) containing 10% Fetal Bovine
567 Serum (FBS) (GIBCO, 10438026). Jar cells (provided by Carolyn Coyne) were cultured
568 in RPMI (GIBCO, 11875093) containing 10% FBS. JEG3 cells were cultured in MEM
569 (GIBCO, 11095080) containing 10% FBS. Culture medium for these cell lines was
570 supplemented with sodium pyruvate (GIBCO, 11360070), glutamax (GIBCO, 35050061),
571 and Penicillin Streptomycin (GIBCO, 15140122) according to manufacturer
572 specifications. H1-ESCs (obtained from WiCell) were grown on Matrigel (Corning,
573 356277) coated plates in MTESR+ (Stemcell, 05825) growth-media and sub-cultured
574 using Accutase (Innovative Cell Technologies, AT-104) and MTESR+ supplemented with
575 CloneR (Stemcell, 05888). All cell lines were cultured at 37°C and 5% CO_2.
576
577 **Vector cloning**
578 DHIV3-GFP, phCMV-RD114env, psi(-)-amphoMLV plasmids were provided by Vicente
579 Planelles (University of Utah). pCGCG-SMRVenv plasmid was provided by Welkin
580 Johnson (Boston University). psPAX2 and pVSVg plasmids were provided by John Lis
581 (Cornell University). The following cloning approaches were performed using primers and
582 constructs described in Supplementary Table 3. HERVH1env ORF was PCR-amplified
583 using Q5 polymerase (NEB, M0491L) from HeLa and 293T genomic DNA respectively
584 and cloned into a TOPO vector (ThermoFisher, 450245).

585 To generate stable *SYN1* and *SUPYN* knock-down cell lines, pHIV lentiviral constructs
586 containing shRNAs targeting *SYN1* and *SUPYN* respectively were cloned using the
587 following strategy. The shRNA encoded in pHIV7-U6-shW3, generously provided by Lars
588 Aagaard (Aarhus University), targets *SYN1*[72]. *SUPYN*-targeting shRNAs were designed
589 using siRNA sequences employed by Jun Sugimoto[4] as a template. pHIV7 lentiviral
590 constructs were cloned using the pHIV7-U6-shW3 plasmid[72] as a template. pHIV7-U6-
591 shSup-cer, pHIV7-U6-shSup-puro, pHIV7-U6-shC-cer, pHIV7-U6-shC-puro, pHIV7-U6-
592 shSyn1-cer, pHIV7-U6-shSyn1-puro were generated using a Gibson assembly approach.
593 To replace the native GFP marker of pHIV7-U6-shW3 with a Cerulean reporter or
594 puromycin resistance marker, we digested pHIV7-U6-shW3 with NheI (NEB, R3131S)
595 and KpnI (NEB, R3142S). This digest resulted in the production of three DNA fragments:
596 pHIV7 backbone, GFP-, and WPRE-containing fragments. We separately PCR amplified
597 each selection marker and WPRE containing pHIV7 fragment. InFusion cloning was then
598 used to ligate the digested pHIV7 backbone to the Cerulean or puromycin cassette and
599 WPRE containing PCR product. shRNAs were cloned into the pHIV7-
600 Cerulean/puromycin transfer construct previously digested with NotI (NEB, R0189S) and
601 NheI. U6-promoter containing shRNA cassettes and the CMV promoter driving marker
602 cassette expression were PCR amplified and subsequently InFusion cloned into the
603 NotI/NheI digested pHIV7-cerulean/puromycin backbone.
604 All pHCMVenv and SUPYN expression constructs, described in this study, were
605 generated as follows: HA-tagged and untagged ORFs with pHCMV homologous
606 overhanging sequence were either PCR amplified using Q5 polymerase (NEB, M0491S)
607 or synthesized (IDT) (see **Supplementary Table 2**), and cloned into EcoRI (NEB,

608 R3101T) digested pHCMV backbones using the InFusion cloning kit (Takara Bio,
609 638920). To generate siRNA-resistant SUPYN rescue constructs, we replaced the native
610 signal peptide sequence[4] (which is targeted by siRNAs used in this study) with (1) a
611 *Gaussia princeps* luciferase SP (SUPYN-lucSP)[73,74] and (2) a shSUPYN resistant
612 SUPYN rescue construct (SUPYN-rescSP) in which the codons were modified to retain
613 the codon identity but disrupt siRNA binding.
614
615 **Antibodies**
616 All antibodies used in this study are commercially available. α-GAPDH (D4C6R,
617 D16H11), α-βactin (D6A8), α-HA (C29F4), α-ASCT2 (V501) primary antibodies were
618 purchased from Cell Signaling Technology. α-Mouse (#7076) and α-Rabbit (#7074) HRP
619 conjugated secondary antibodies were purchased from Cell Signaling Technology. IRDye
620 secondary antibodies were purchased from Licor (925-32211, 925-68072, 925-32210,
621 925-68073). α-SUPYN primary antibody was purchased from Phoenix Pharma (H-059-
622 052). Alexa-fluor conjugated secondary antibody was purchased from Invitrogen.
623
624 **Western Blot**
625 Whole cell extracts from cultured cell lines were prepared using 1x GLO lysis buffer
626 (Promega, E266A). One third volume of 4x Laemli buffer was added to one volume whole
627 cell extract samples, then incubated at 95°C for 5 minutes, and sonicated for 15 minutes
628 at 4°C (amplitude 100; pulse interval 15 seconds on, 15 seconds off). Approximately 30
629 ug of protein were separated by SDS-PAGE (BioRad, 1610175), transferred to PVDF
630 membrane (BioRad, 1620177), blocked according to antibody manufacturers

631 specification, and incubated overnight in appropriate primary antibody then incubated in
632 IRDye or peroxidase conjugated goat anti-mouse or anti-rabbit secondary antibodies for
633 1 hour at room temperature. Protein was then detected using ECL reagent (BioRad,
634 1705061) or the Licor Odyssey imaging system.
635
636 **IF microscopy**
637 Human second trimester placental tissue that resulted from elective terminations was obtained
638 from the University of Pittsburgh Health Sciences Tissue Bank through an honest broker system
639 after approval from the University of Pittsburgh Institutional Review Board (IRB) and in
640 accordance with the University of Pittsburgh's tissue procurement guidelines. Tissue was
641 excluded in cases of fetal anomalies or aneuploidy. Third trimester placental tissue was obtained
642 through the Magee Obstetric Maternal & Infant (MOMI) Database and Biobank after approval from
643 the University of Pittsburgh IRB. Women who had previously consented for tissue donation and
644 underwent cesarean delivery were included. Placental tissues were fixed in 4% PFA (in 1x
645 PBS) for 30 minutes, permeabilized with 0.25% Triton X-100 for 30 minutes (on a rocker),
646 washed with 1x PBS and then incubated with primary anti-Suppressyn antibody at 1:200
647 in 1x PBS for 2-4 hours at room temperature. These samples were incubated with Alexa-
648 fluor conjugated secondary antibody (Invitrogen) diluted 1:1000 and counterstained with
649 actin. DAPI was included in our PBS and then mounted in Vectashield mounting medium
650 with DAPI (Vector Laboratories, H-1200).
651
652 **Virus production**
653 Low passage 293T cells were used to produce lentiviral particles. DHIV3-GFP and env-
654 expression plasmids were co-transfected at a mass ratio of 2:1 using lipofectamine 2000

(ThermoFisher, 11668030). shRNA encoding lentiviral particles were produced by co-transfecting pHIV7, psPAX2, pVSVg according to BROAD institute lentiviral production protocol (https://portals.broadinstitute.org/gpp/public/resources/protocols) using Lipofectamine 2000. Growth media was replaced on transfected cells after overnight incubation. At 72 hours post-transfection, virus containing supernatant was harvested, centrifuged to remove cell debris, filtered through a 0.45 um pore filter, and stored at -80°C.

Infection Assays

293T cells were transfected with env-overexpression constructs using Lipofectamine 2000 and incubated 24 hours. Transfected cells were infected with reporter virus by applying virus (HIV-RD114env, HIV-VSVg, HIV-SMRVenv) stocks in the presence of polybrene (Santa Cruz Bio, sc-134220) at a final concentration of 4 ug/mL. After 6-8 hours, virus stock was replaced with fresh growth media. Infected cells were maintained for 72 hours, replacing media when necessary, and harvested with trypsin. Detached cells were suspended in fresh growth media, strained and analyzed by flow cytometry. For the H1-ESC infection experiment, relative infection rates were calculated by normalizing the percent GFP+, HIV-RD114env infected cells to the percent GFP+ HIV-VSVg infected cells. For env/SUPYN overexpression experiments, relative infection rates were calculated by normalizing the percent GFP+ env/SUPYN-transfected cells to the percent GFP+ empty vector transfected cells. ANOVA with Tukey HSD tests were implemented in R (v3.6.3).

678 **Placental cell shRNA transduction**

679 Placenta-derived cell lines were treated with pHIV-shRNA-virus-containing supernatant
680 and incubated for 72 hours as described in Infection Assays. Cerulean positive cells were
681 sorted using the BD FACS Aria cytometer. Cells transduced with puroR cassette were
682 treated with Puromycin (GIBCO, A1113802) at a final concentration of 3.5 ug/mL for 7
683 days, then cultured in regular growth media.

684

685 **RT-qPCR**

686 RNA was isolated from cultured cells using the RNeasy Mini Kit (Qiagen, 74104) and an
687 on column dsDNAse digestion (Qiagen, 79254) was performed. 1-3 ug of total RNA were
688 used to generate cDNA with the maxima cDNA synthesis with dsDNAse kit
689 (ThermoFisher, K1681). qPCR reactions were performed using the LC480 Instrument
690 with Sybr Green PCR master mix (Roche, 04707516001) according to manufacturer's
691 protocol and using primers indicated in Supplementary Table 2. Gene expression was
692 then quantified using the $\Delta\Delta$CT method[75]. 18S expression was used as a reference
693 housekeeping gene. Wilcox rank sum tests were performed using R (v3.6.3).

694

695 **Envelope evolutionary sequence analyses**

696 Orthologous *SUPYN*, *SYN1*, and *SYN2* sequences were extracted from the 30-species
697 MULTIZ alignment[14] and formatted for sequence alignment using the phast package[76].
698 These and additional syntenic *SUPYN* and *SYN2* open reading frame sequences were
699 validated/identified by BLASTn[77] search with default settings of publicly available
700 Catarrhine primate genomes (ncbi.nih.gov). Mariam Okhvat of the Carbone Lab (Oregon

Health and Science University) generously provided BAM files containing read alignment information for *SUPYN*, *SYN1*, and *SYN2* generated from whole genome sequencing of *Hoolock leuconedys* (Hoolock Gibbon), *Symphalangus syndactylus* (Siamang), *Hylobates muelleri* (Müller's Gibbon), *Hylobates lar* (Lar Gibbon), *Hylobates moloch* (Silvery Gibbon), *Hylobates pileatus* (Pileated Gibbon), and *Nomascus gabriellae* (Yellow-cheeked Gibbon). Where multiple individuals were sequenced, a consensus sequence was generated using samtools[67] and JalView[78].

To perform *dN/dS* analyses, orthologous env sequences (>90bp length) encoding the mature sequence downstream of the signal peptide cleavage site, were aligned using MEGA7[79] and manually converted to PHYLIP format. A Newick tree was generated based on this alignment using the maximum likelihood algorithm implemented in MEGA7. *Codeml*, implemented in the PAML package, was run to calculate *dN/dS* values and log likelihood (LnL) scores generated under models M0, M1, M2, M7 and M8[38]. Chi-square tests comparing LnL scores generated under models of neutral evolution and selection were performed.

We used two approaches to reconstruct ancestral hominoid and OWM *SUPYN* sequences. First, we reconstructed ancestral *SUPYN* sequences using the majority rule consensus sequence (calculated in JalView) of the hominoid and OWM clade respectively. At positions where nucleotide identity was ambiguous, the dominant nucleotide identity in the neighboring clade was used as a tiebreaker. These sequences were used for our infection assays shown in **Fig 3c.** We also employed a maximum likelihood approach using the *baseml* program, implemented in PAML[38]. We reconstructed ancestral *SUPYN* sequences using the hominoid species, shown in **Fig 3a**,

724 and the 6 OWM monkeys with the most complete *SUPYN*-coding open reading frame
725 (olive baboon, drill, crab-eating macaque, rhesus macaque, japanese macaque, green
726 monkey) as our input sequences. Because PAML requires a Newick tree as an input, the
727 MEGAX[80-82] maximum likelihood algorithm was used to generate a Newick tree with the
728 above described *SUPYN* sequences. *Baseml* was run using models 3-7 (F84, HKY85,
729 T92, TN93, REV). As shown in **Extended Data 10**, both the consensus- and maximum
730 likelihood reconstructions were identical for the OWM *SUPYN* sequences. The
731 consnensus-based hominoid sequence reconstruction differed from our maximum
732 likelihood-based reconstruction by two amino acids. These two positions are unlikely to
733 affect the function of the resulting protein because these sites are identical to siamang
734 *SUPYN*, which restricts RD114env-mediated infection **(Fig 3b, c)**.
735
736 **Genome-wide search for endogenous retrovirus derived envelope open reading**
737 **frames**
738 Candidate envelope open reading frames (envORF) were identified by performing
739 tBLASTn[83] searches of the hg19 human genome assembly using envelope amino acid
740 sequences, obtained from Repbase[84] and published retroviral envelope sequences, as a
741 query. Collected hits were used as a query to repeat a tBLASTn search, initially yielding
742 82715 candidate envORF sequences. This list of candidates was filtered using the
743 following criteria. (1) EnvORFs must have a length of ≥70 aa. (2) Hits starting at position
744 ≥300 aa were removed because such open reading frames are predicted to encode a
745 portion of the envelope transmembrane domain, which is not expected to play a role in
746 receptor binding. (3) After these processing steps, our list was further concatenated to

only include unique open reading frame sequences (n=1507) in the hg19 genome assembly.

To estimate envORF expression, we mined publicly available scRNA-seq datasets generated on the SMART-seq2 platform because this technology yields higher genomic coverage compared with methods employing poly(A) enrichment. We gathered scRNA-seq datasets in fastq format from preimplantation embryos[47] (GSE36552), placenta[48] (GSE89497), primordial germ cells[85] (GSE63818), hematopoiesis[86] (GSE75478), neuronal differentiation[87] (GSE93593), prefrontal cortex[88] (GSE67835), and pancreas[89] (GSE81547) (Supplemental Table 1). We also included bulk RNA-seq of immune cells from six individuals comprising 25 subtypes[58] (GSE118165) (Supplemental Table 1).

We curated our envORF sequences (in fasta format) and hg19 refseq gene models[14] (gtf format) to calculate envORF expression with high confidence. First, we masked envORF DNA sequences from the hg19 genome. These sequences were added, as individual fasta contigs, to the hg19 reference assembly. We then modified the transcriptome model by appending the envORFs models (gtf format) to the hg19 refseq gene models. Next, we indexed the modified reference sequences using STAR default parameters, by providing the new transcriptome models that include the envORFs. This approach enabled us to simultaneously calculate envORF and protein-coding gene expression. After aligning and retaining uniquely mappable sequencing reads, we calculated envORF and gene expression (see Single cell RNA-seq analysis procedures above). We processed the individual datasets independently, then constructed the expression matrix using *Seurat* and checked if the obtained clusters were consistent with the original studies. Despite including the envORFs, we identified each cell type in similar proportions,

as shown in their corresponding studies (Extended Data 11). Then we merged all of the datasets into a single pool for our downstream analysis while maintaining cell type and tissue annotations. We scaled and normalized these datasets using Seurat within the R environment and checked the normalization status of merged datasets with a UMAP biplot. This was done by using the most variable genes to ensure that cells did not cluster based on platform or batch differences (Extended Data 11 and Supplemental Table 1). We calculated gene and envORF expression (TPM) using uniquely mapped reads. Any envORF with TPM < 1 in > 99% of cells across the data frame was discarded from further analysis. This strategy identified 668 unique envORFs with evidence for expression. Finally, expression of each envORF across all cell types was used to cluster envORF according to their expression profile, which were visualized in Fig. 4.

To survey the expression of known envelope genes, we utilized Gtex[90] bulk RNA-seq datasets (phs000424.v6.p1), which were generated from distinct post-mortem tissues.

Methods References

47. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20,** 1131–1139 (2013).

48. Liu, Y. *et al.* Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta. *Cell Res.* **28,** 819–832 (2018).

49. Pavličev, M. *et al.* Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* **27,** 349–361 (2017).

50. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).

52. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177,** 1888–1902.e21 (2019).

53. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14,** 309–315 (2017).

54. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563,** 347–353 (2018).

55. Zadora, J. *et al.* Disturbed Placental Imprinting in Preeclampsia Leads to Altered Expression of DLX5, a Human-Specific Early Trophoblast Marker. *Circulation* **136,** 1824–1839 (2017).

56. D Antonio, M. *et al.* Identifying DNase I hypersensitive sites as driver distal regulatory elements in breast cancer. *Nat Commun* **8,** 436 (2017).

57. Chung, H. *et al.* Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. *Cell* **172,** 811–824.e14 (2018).

58. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Publishing Group* **51,** 1494–1505 (2019).

59. Tsankov, A. M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518,** 344–349 (2015).

814 60. Barakat, T. S. *et al.* Functional Dissection of the Enhancer Repertoire in Human
815 Embryonic Stem Cells. *Cell Stem Cell* **23,** 276–288.e8 (2018).
816 61. Kwak, Y.-T., Muralimanoharan, S., Gogate, A. A. & Mendelson, C. R. Human
817 Trophoblast Differentiation Is Associated With Profound Gene Regulatory and
818 Epigenetic Changes. *Endocrinology* **160,** 2189–2203 (2019).
819 62. Dunn-Fletcher, C. E. *et al.* Anthropoid primate-specific retroviral element THE1B
820 controls expression of CRH in placenta and alters gestation length. *PLoS Biol.* **16,**
821 e2006337 (2018).
822 63. Krendl, C. *et al.* GATA2/3-TFAP2A/C transcription factor network couples human
823 pluripotent stem cell differentiation to trophectoderm with repression of
824 pluripotency. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E9579–E9588 (2017).
825 64. Gao, L. *et al.* Chromatin Accessibility Landscape in Human Early Embryos and Its
826 Association with Evolution. *Cell* **173,** 248–259.e15 (2018).
827 65. Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic
828 transition during ZGA. *Nature* **557,** 256–260 (2018).
829 66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*
830 *Methods* **9,** 357–359 (2012).
831 67. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
832 **25,** 2078–2079 (2009).
833 68. Gasper, J.M. Improved peak-calling with MACS2. *biorxiv.org* doi:
834 https://doi.org/10.1101/496521 (2018)
835 69. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification
836 of Problematic Regions of the Genome. *Sci Rep* **9,** 9354 (2019).

837 70. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
838 genomic features. *Bioinformatics* **26,** 841–842 (2010).
839 71. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnology* **29**, 24-26
840 (2011).
841 72. Aagaard, L. *et al.* Silencing of endogenous envelope genes in human
842 choriocarcinoma cells shows that envPb1 is involved in heterotypic cell fusions.
843 *J. Gen. Virol.* **93,** 1696–1699 (2012).
844 73. Luft, C. *et al.* Application of Gaussia luciferase in bicistronic and non-conventional
845 secretion reporter constructs. *BMC Biochem.* **15,** 14 (2014).
846 74. Knappskog, S. *et al.* The level of synthesis and secretion of Gaussia princeps
847 luciferase in transfected CHO cells is heavily dependent on the choice of signal
848 peptide. *J. Biotechnol.* **128,** 705–715 (2007).
849 75. Schmittgen, T. D. Livak, J. L. Analyzing real-time PCR data by the comparative
850 CT method. *Nat. Protocols* **3**, 1101-1108 (2008).
851 76. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic
852 analysis with space/time models. *Brief. Bioinformatics* **12,** 41–51 (2011).
853 77. Ye, J., McGinnis, S. & Madden, T. L. BLAST: improvements for better sequence
854 analysis. *Nucleic Acids Res.* **34,** W6–9 (2006).
855 78. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J.
856 Jalview Version 2--a multiple sequence alignment editor and analysis workbench.
857 *Bioinformatics* **25,** 1189–1191 (2009).

858 79. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics
859 Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33,**
860 1870–1874 (2016).
861 80. Tamura, K., Nei, M. Estimation of the number of nucleotide substitutions in the
862 control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*.
863 **10(3)**, 512-526 (1993).
864 81. Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. MEGA X: molecular
865 evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* **35(6),**
866 1547-1549 (2018).
867 82. Stecher, G. *et al.* Molecular evolutionary genetics analysis (MEGA) for macOS.
868 *Mol Biol Evol*. **37(4)**, 1237-1239 (2020).
869 83. Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A. & Altschul, S. F.
870 Composition-based statistics and translated nucleotide searches: improving the
871 TBLASTN module of BLAST. *BMC Biol.* **4,** 41 (2006).
872 84. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive
873 elements in eukaryotic genomes. *Mob DNA* **6,** 11–6 (2015).
874 85. Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human
875 Primordial Germ Cells. *Cell* **161,** 1437–1452 (2015).
876 86. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a
877 continuous process. *Nat. Cell Biol.* **19,** 271–281 (2017).
878 87. Close, J. L. *et al.* Single-Cell Profiling of an In Vitro Model of Human Interneuron
879 Development Reveals Temporal Dynamics of Cell Type Production and
880 Maturation. *Neuron* **93,** 1035–1048.e5 (2017).

881 88. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single
882 cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 7285–7290 (2015).
883 89. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional
884 Signatures of Aging and Somatic Mutation Patterns. *Cell* **171,** 321–330.e14
885 (2017).
886 90. GTEx Consortium. *et al.* Genetic effects on gene expression across human
887 tissues. *Nature* **550,** 204–213 (2017).

888

889 **Acknowledgements**

890 This work was funded by Cornell University and by the National Institutes of Health to CF
891 (R01 GM112972, R35 GM122550). We thank Dr. Vicente Planelles for sharing reporter
892 virus plasmids and training; Dr. Welkin Johnson for sharing SMRVenv expression
893 plasmid; Dr. Lars Aagaard for sharing shRNA transduction constructs; Dr. John Lis for
894 providing lentiviral packaging construct; Dr. Amnon Koren and Rita Rebello for providing
895 embryonic stem cell cultures and technical support respectively**;** Dr Lucia Carbone and
896 Dr. Mariam Okhvat for sharing unpublished gibbon genome sequences. We thank Ray
897 Malfavon-Borja for producing an initial list of envelope open reading frames in the human
898 genome. We thank Maia Clare for her contribution to evolutionary sequence analyses.
899 We thank members of the Feschotte lab for helpful advice and discussion throughout the
900 project.

901

902

903

Author contributions

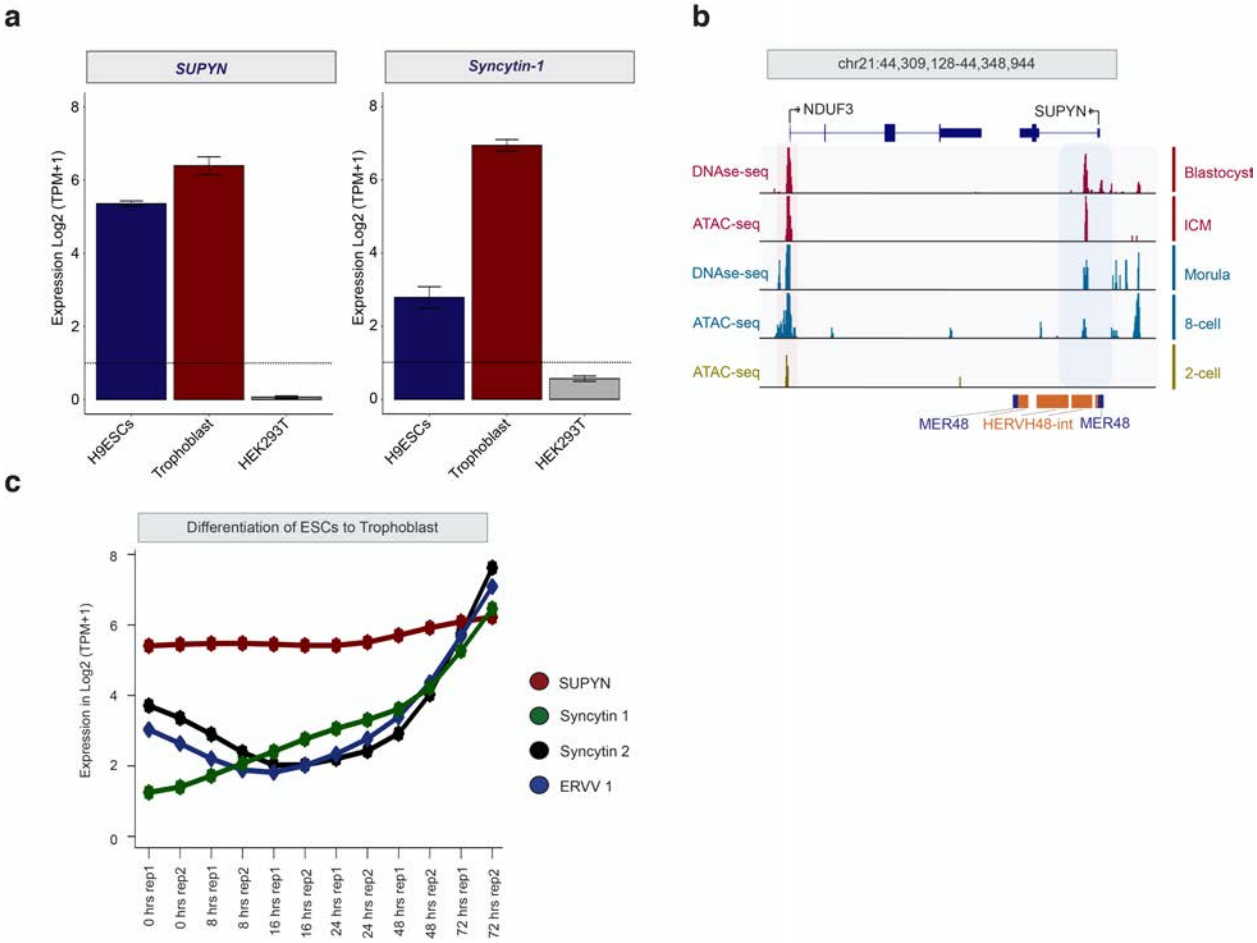
CF conceived of this project. JAF and CF designed and developed this project. JAF designed and conducted all experiments, evolutionary sequence analyses, and analyzed all experimental data. MS performed all gene expression and regulation analyses. HBC and RAK helped perform infection assays and evolutionary sequence analyses. CBC performed placental stains. JAF, CF and MS wrote this manuscript.

Competing interest declaration

The authors declare no competing interests.

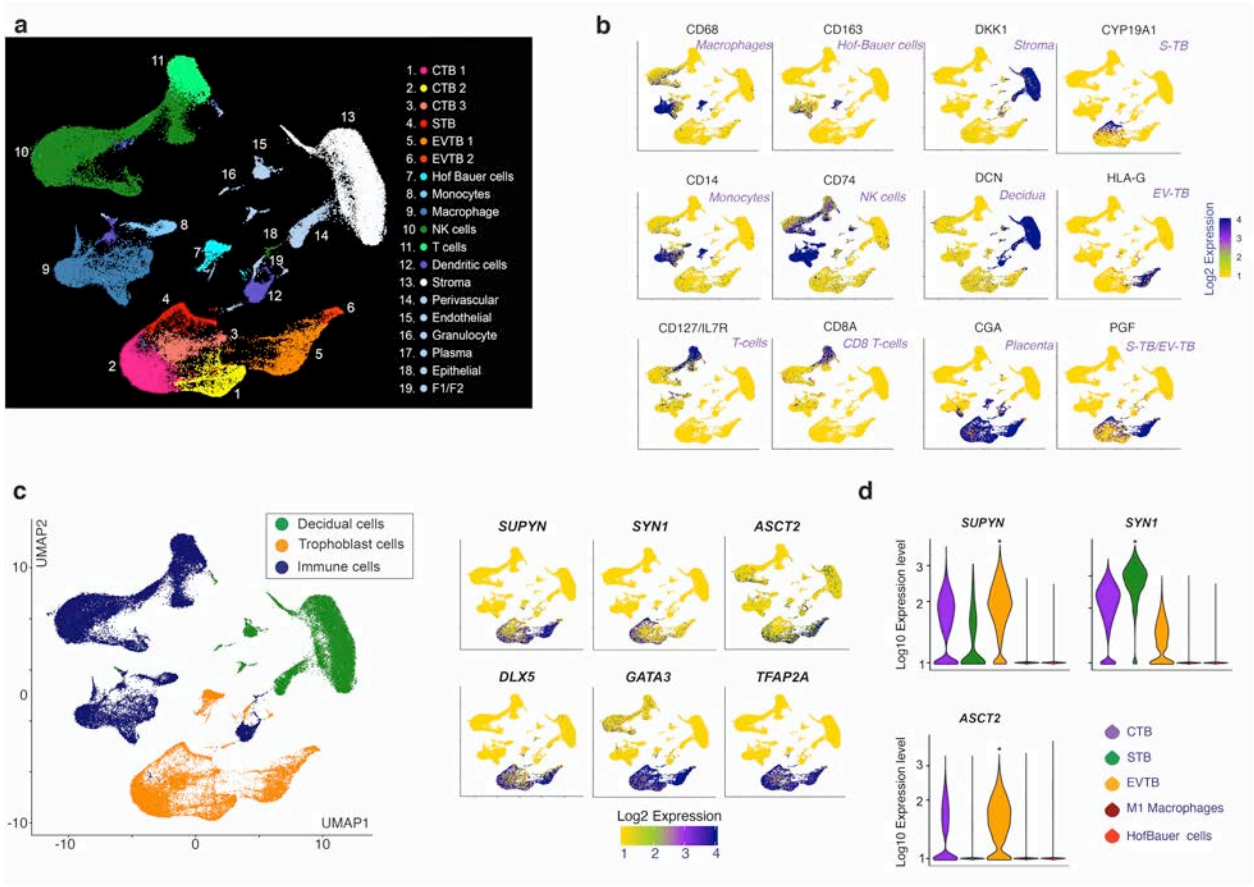
Additional Information

Extended Data



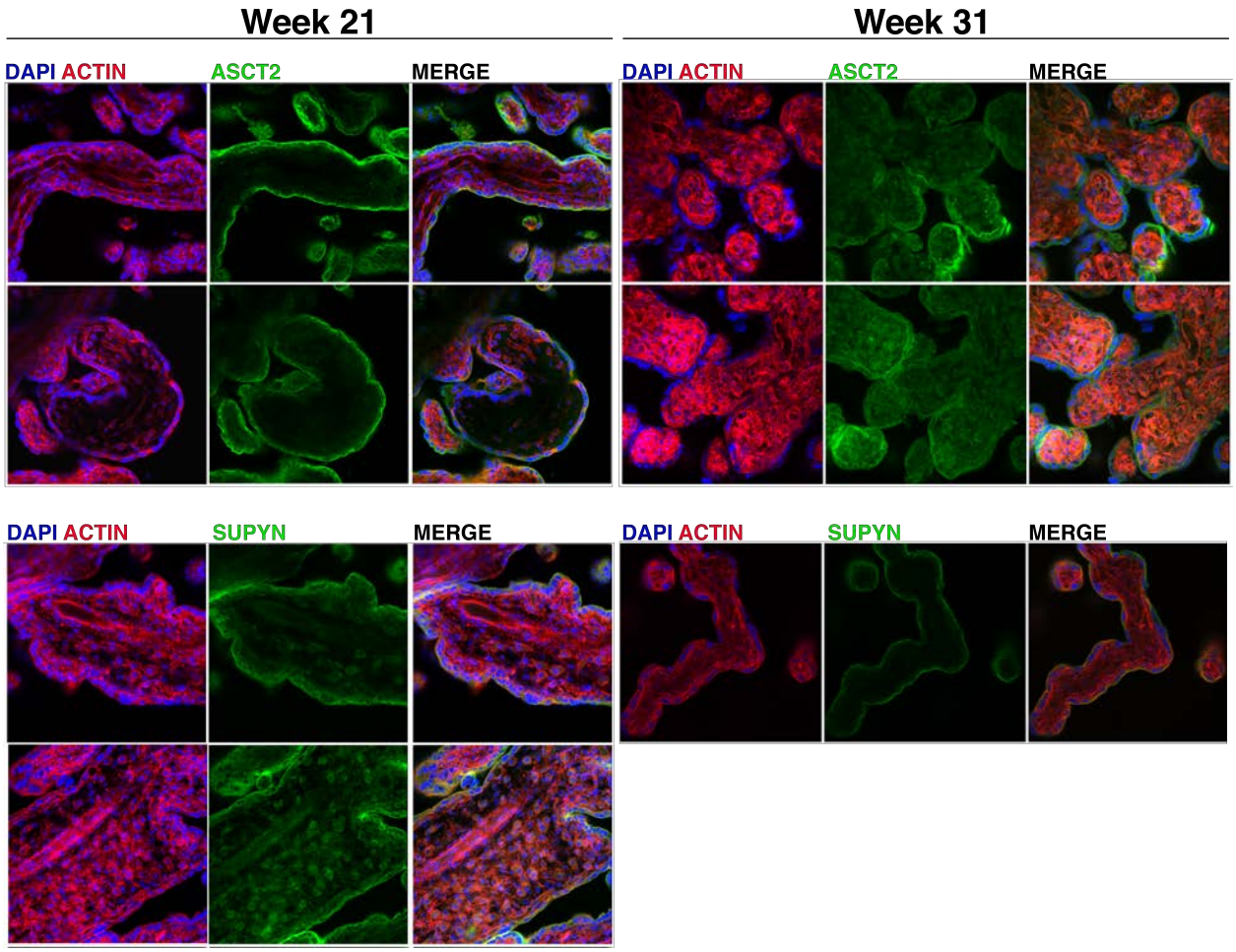
Extended Data Figure 1: SUPYN is expressed in human early embryo and placenta but not in 293T cells.

(a) Bar graphs show SUPYN and SYN1 expression in indicated cell types. **(b)** Genome browser view showing ATAC-seq and DNAse-seq signals at the SUPYN locus, including upstream and downstream sequences. Framed region highlights overlapping peaks at the SUPYN locus. **(c)** Line plot depicts HERVenv gene expression level during BMP4-mediated *in vitro* hESCs to trophoblast differentiation. Time points correspond to cells harvested 8hr, 24hr, 48hr, and 72hr post BMP4 treatment.

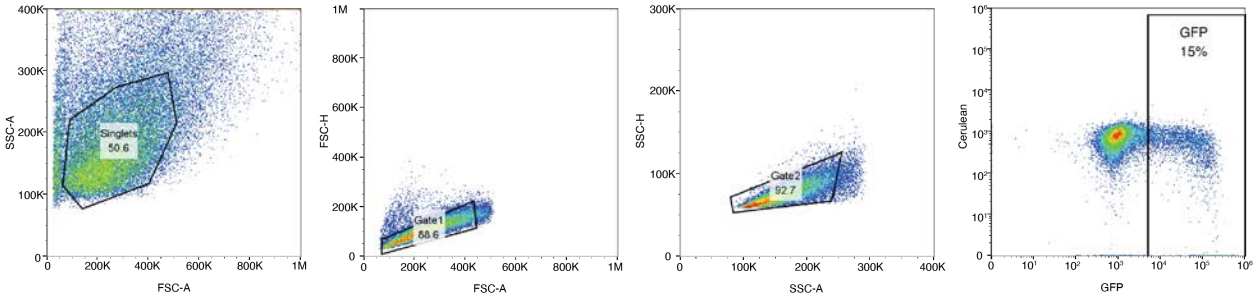


Extended Data Figure 2: Defining lineage-specific SYN1, SUPYN, and ASCT2 expression from placental single-cell transcriptomics

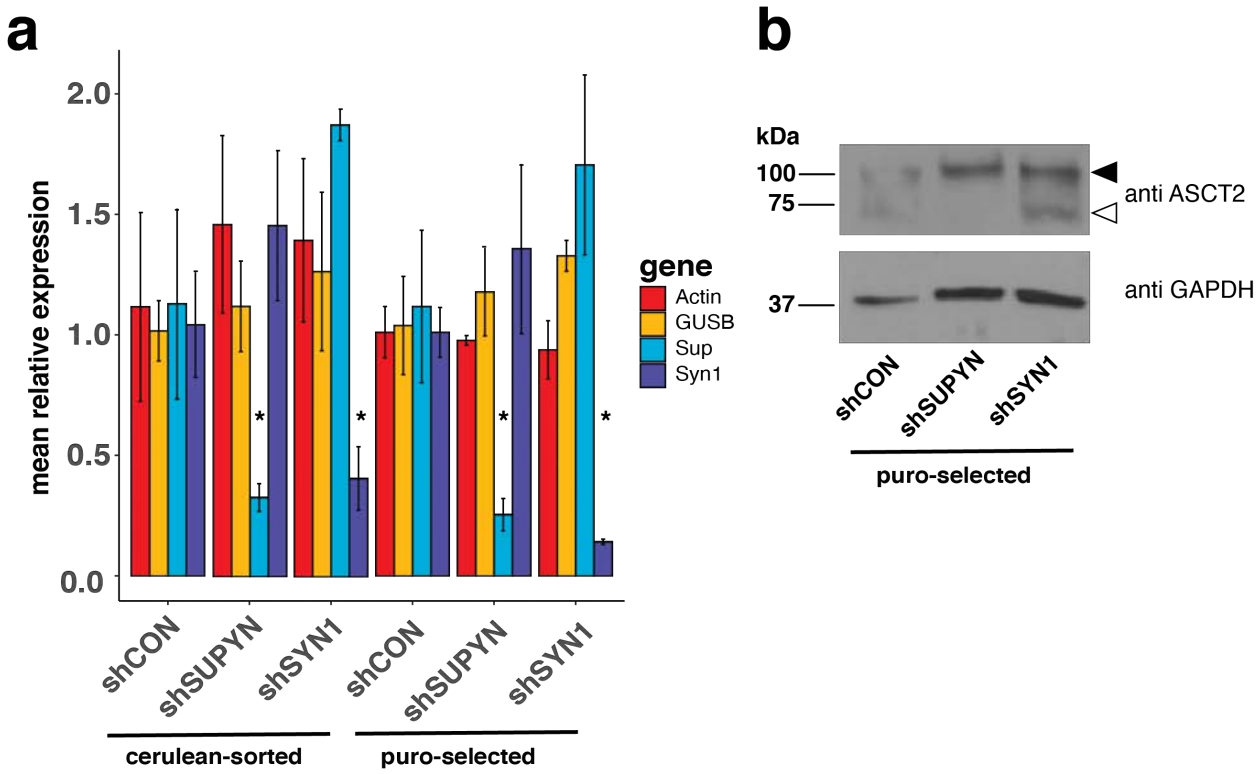
(a) UMAP plot generated from published scRNA-seq data generated from 1st trimester placental explants. Colors denote placental (pink, red, orange, and yellow) immune (blue and green) and maternal cell lineages (white and grey). **(b)** Feature plots visualize single-cell expression level of lineage-defining marker genes. **(c)** Simplified UMAP plot, shown in **a,** of scRNAseq data displaying trophoblast (yellow), decidual (green) and immune (purple) cell identity. Sub-panels display single-cell-level expression of indicated genes. **(d)** Violin plots denote single-cell *SUPYN* and *ASCT2* expression in multiple placental-cell lineages.



Extended Data Figure 3: ASCT2 and SUPYN expression in 2nd and 3rd trimester human placenta. Confocal microscopy of 2nd (week 21) and 3rd (week 31) trimester placental villi explants. Villi were stained for ASCT2 (green upper panels) or SUPYN (green lower panels) and Actin (red). Cell nuclei are marked with DAPI (blue).

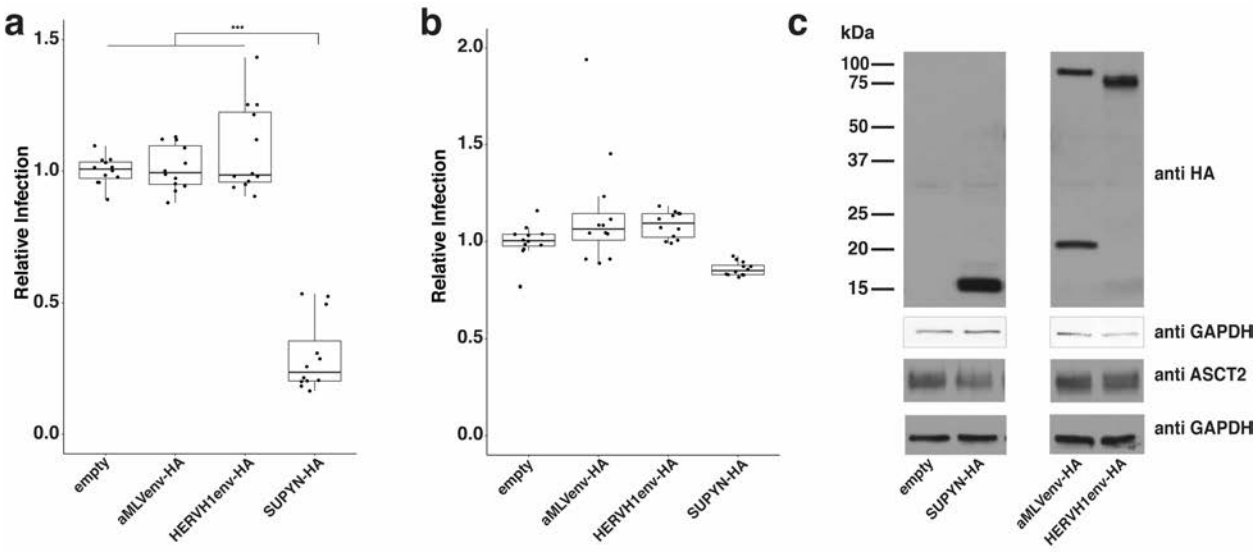


Extended Data Figure 4: Flow Cytometry analysis scheme. Representative sequential gating scheme to assess reporter virus infection rate.



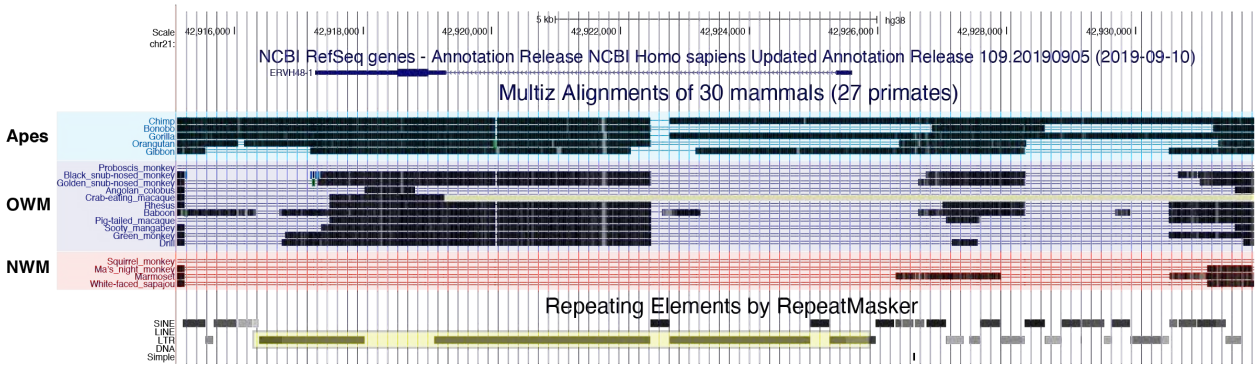
Extended Data Figure 5: Characterization of shRNA transduced Jar cells and validation of env overexpression constructs.

(a, b) *SUPYN* and *SYN1* knock down was validated by qPCR. Bar plots represent mean relative gene expression normalized to shCON in cerulean-sorted and puromycin-selected cell lines respectively (n = 3). Error bars represent ± standard error mean (*$p < 0.1$; Wilcox rank sum test). **(b)** Western Blot analysis (αGAPDH, αASCT2) of shRNA-transduced Jar cell lysates. Putatively glycosylated and unglycosylated ASCT2 are marked by filled and empty arrowheads respectively.



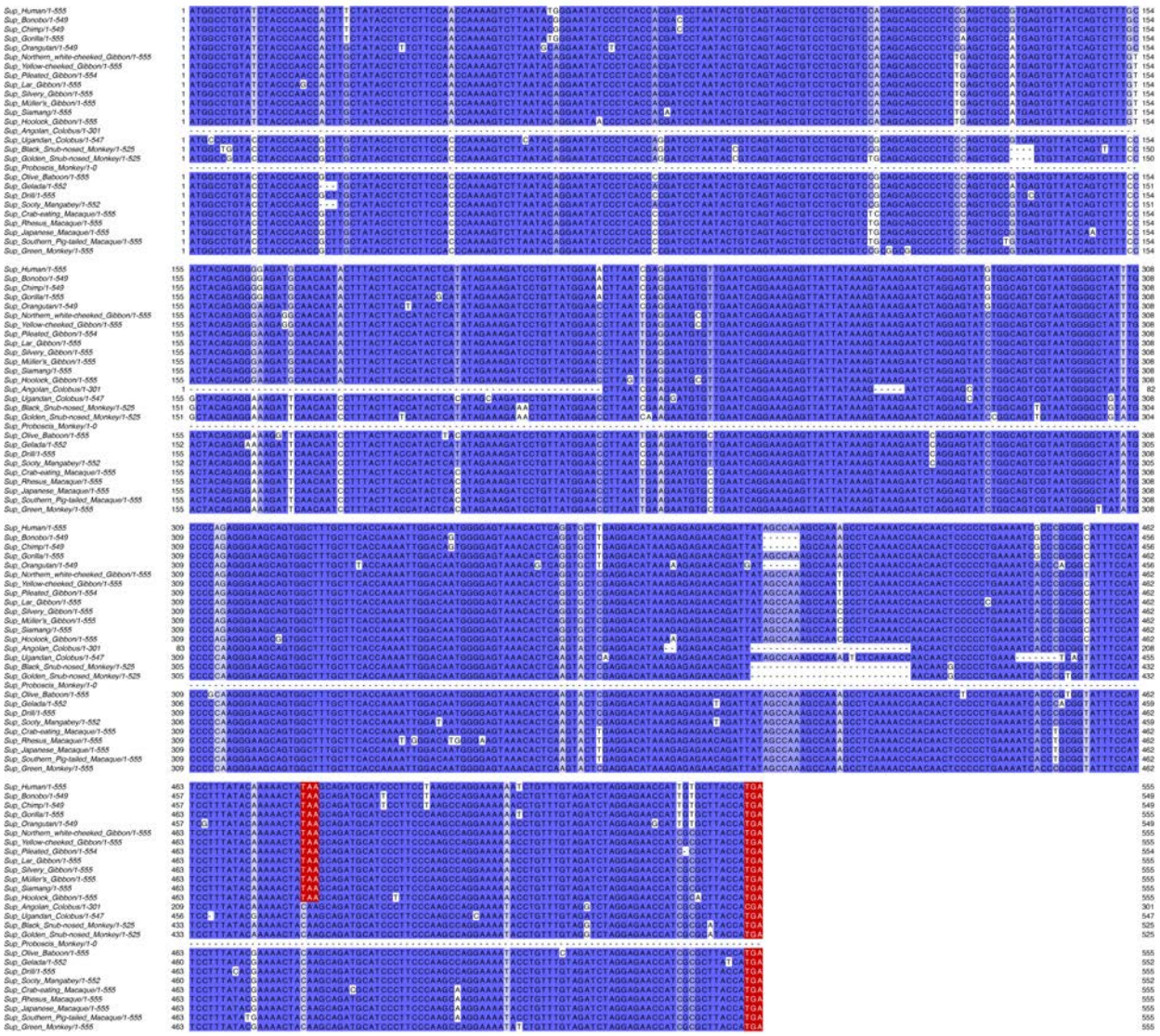
Extended Data Figure 6: SUPYN expression is sufficient to specifically restrict RDR env-mediated infection

(a, b) 293T cells, transfected with HA-tagged SUPYN and env constructs, were infected with HIV-RD114env **(a)** and -VSVg **(b)** respectively. Relative infection rates were determined by normalizing GFP+ counts to empty vector. ($n \geq 3$ with ≥ 1 technical replicate; ***adj. $p < 0.001$; *adj. $p < 0.05$; Tukey HSD). **(c)** Western Blot analysis (αHA, αGAPDH, αASCT2) of 293T cell lysates following transfection with indicated constructs.



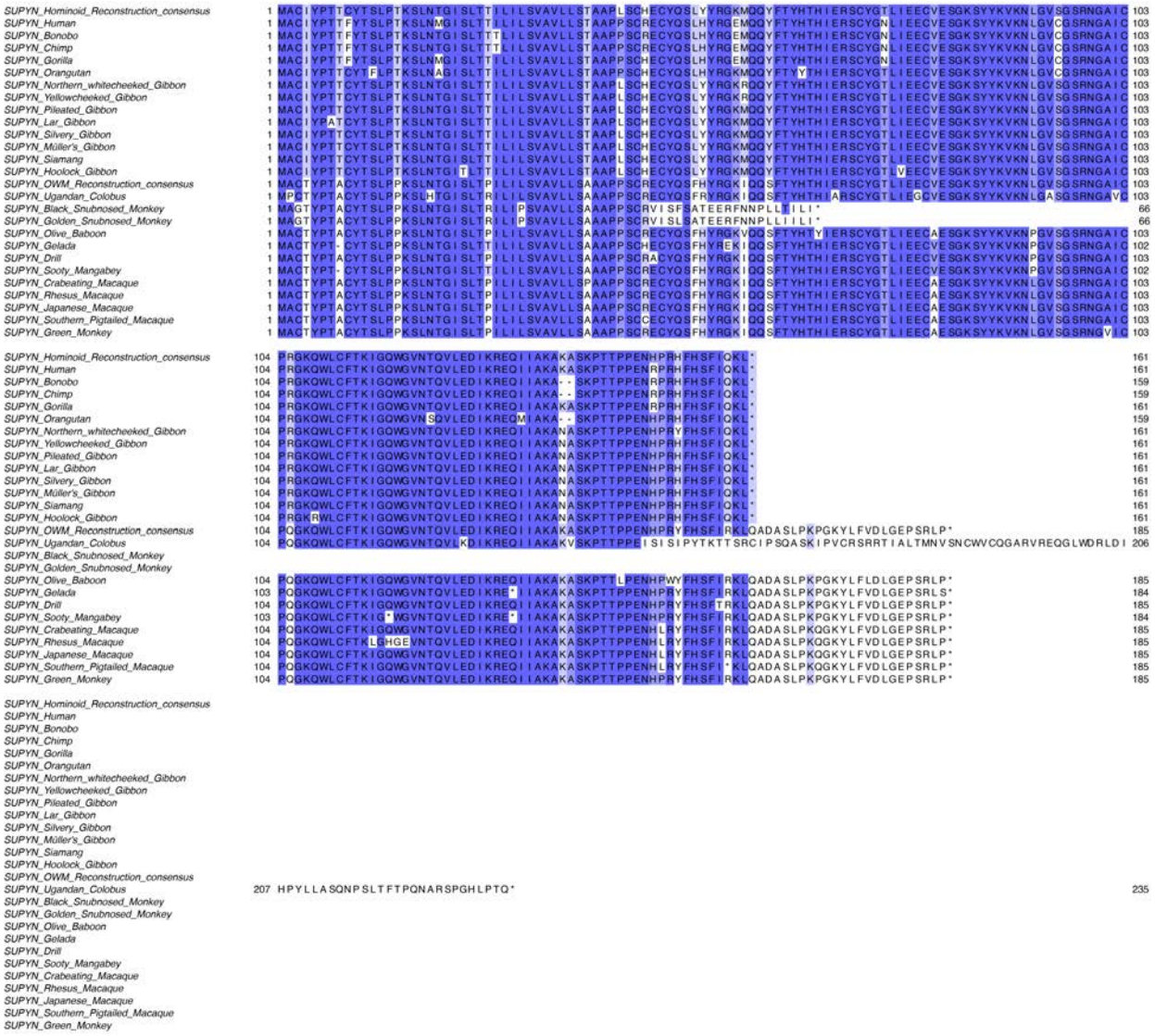
Extended Data Figure 7: SUPYN locus conservation in primates.

UCSC genome Browser snapshot of *SUPYN*-coding locus with surrounding sequence. Modified NCBI RefSeq gene, simian whole genome alignment (from Multiz 30-species track), and RepeatMasker repetitive element tracks are shown. The SUPYN-coding HERVH48 provirus is highlighted by the yellow box.



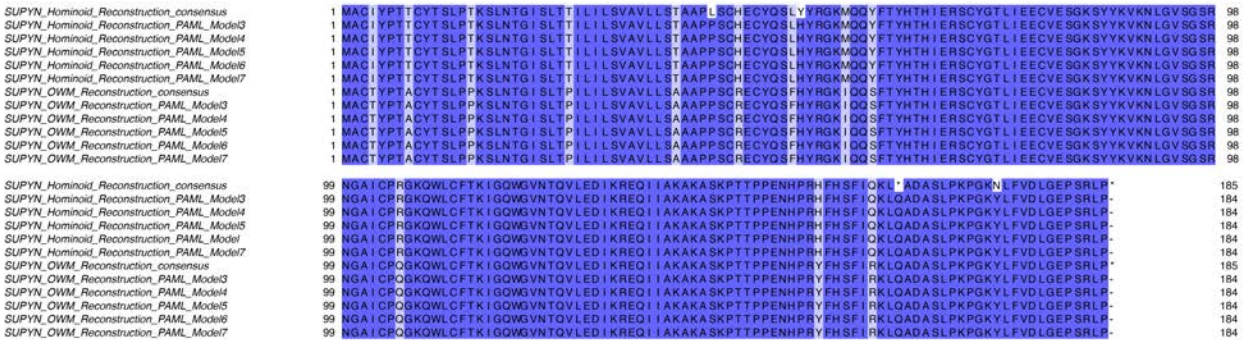
Extended Data Figure 8: Nucleic acid sequence alignment of primate Suppressyn orthologs.

Suppressyn encoding nucleotide sequences are shaded blue based on a minimum sequence identity threshold of 45% (light), 75% (medium) and 80% (dark). Conserved ape-specific and ancestral stop codons are highlighted in red.



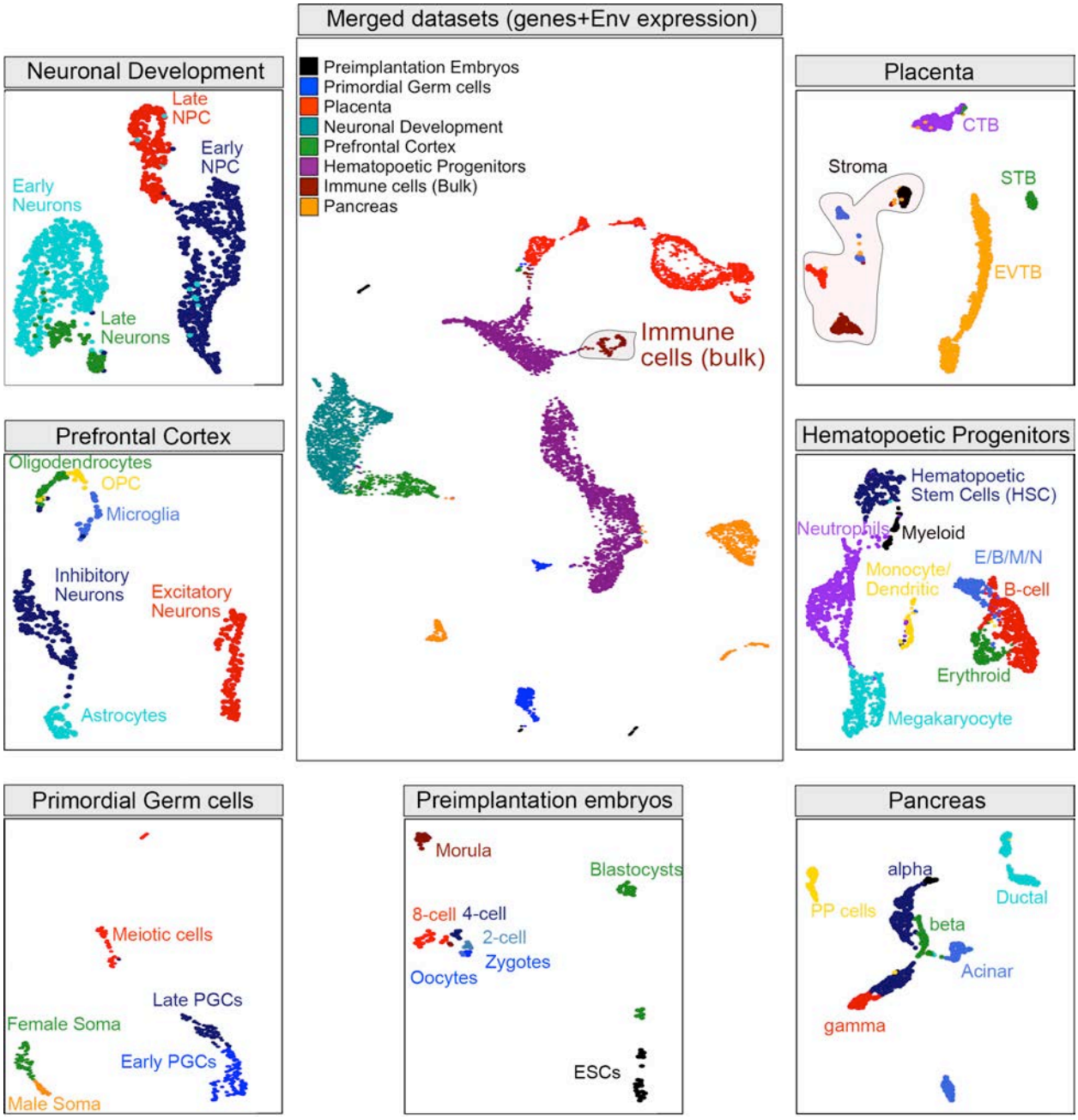
Extended Data Figure 9: Amino Acid sequence alignment of primate SUPYN orthologs.

Primate and ancestral SUPYN peptide sequences are shown. Sequences are shaded blue based on a minimum sequence identity threshold of 45% (light), 75% (medium) and 80% (dark) Ancestral SUPYN sequences are based on our consensus-based sequence reconstruction (see Methods).



Extended Data Figure 10: Amino Acid sequence alignment of ancestral SUPYN sequences.

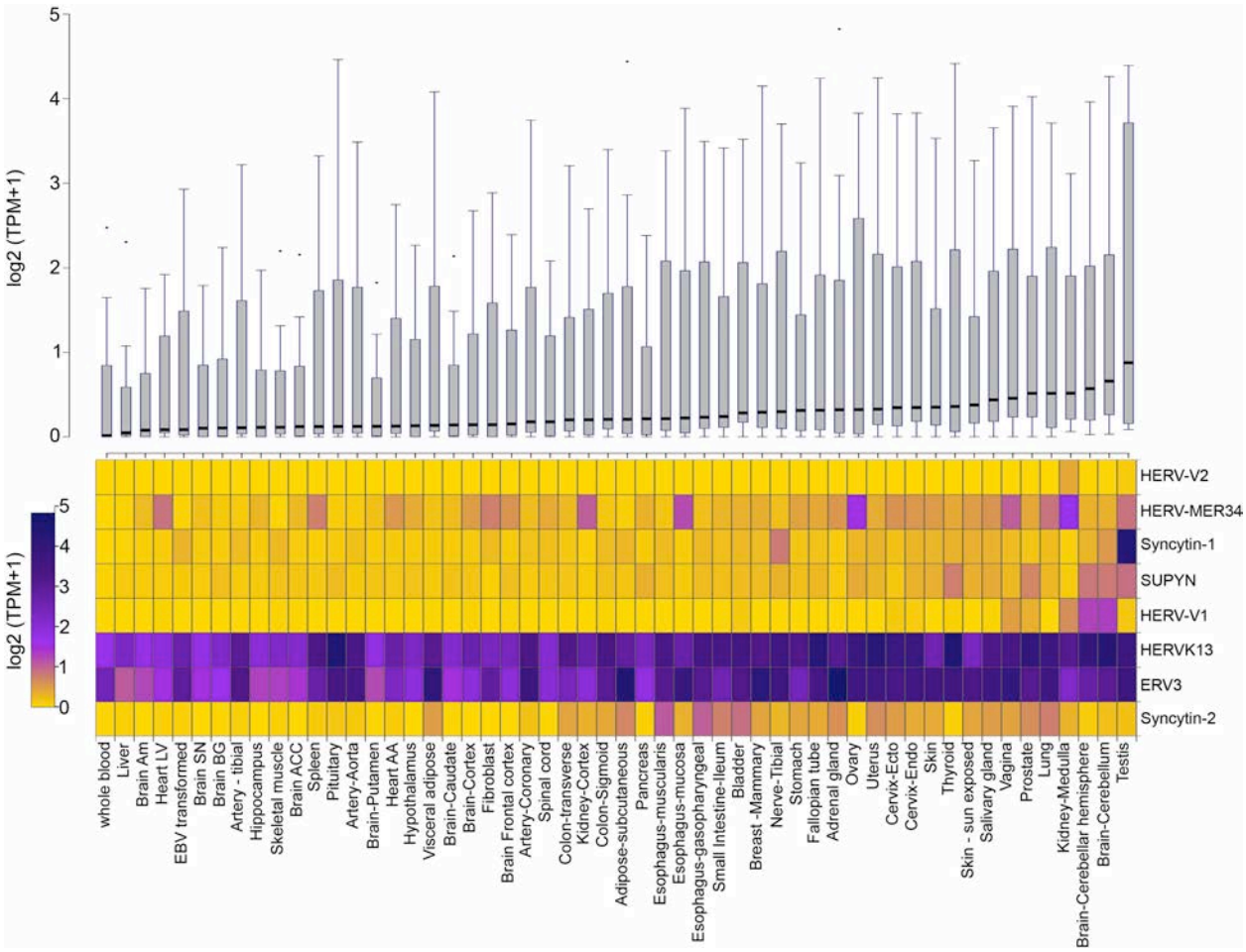
Peptide sequences of consensus and maximum likelihood-based SUPYN sequence reconstructions (see methods) are aligned.



Extended Data Figure 11: Analysis of envORF expression in single cell RNA-seq taken from various human tissue sources

(a) UMAP plots illustrate the Louvain clustering of seven independent sc-RNA-seq datasets corresponding to the development of human embryos and somatic tissues (see Methods). The large central UMAP plot represents our integrative analysis of seven scRNA-seq and one bulk

RNA-seq datasets obtained from independent studies (Supplemental Table 1). Surrounding UMAP plots represent the combined expression of human RefSeq and envORF genes (supplemental Table 3) with cell identities labeled. **(b)** Multiple violin plots demonstrate the expression of annotated ERV envelopes in each dataset shown in **a**.



Extended Data 12: Annotated human endogenous retrovirus envelope expression in various human tissues.

(a) The boxplot shows the transcript expression distribution of annotated ERV envelopes in tissues assayed by the GTEx project (Supplementary Table 1). **(b)** The heatmap displays the

1040 expression (log2 TPM) of individual ERV envelope genes (rows) in each tissue (columns). The color

1041 scheme ranges gradually from no expression (gold) to higher expression (midnight blue).

1042

1043

1044 **Tables**

1045 **Supplementary Table 1: External data sources**

Description	Author	Year	Publication	Methods Reference	Dataset	Tissue source	Figure
scRNAseq	Yan et al.	2013	PMID: 23934149	46	GSE36552	embryo	1a, 4d, Ext 11
	Vento-Tormo et al.	2018	PMID: 30429548	53	E-MTAB-6701	placenta	1d-f, Ext 2
	Liu et al.	2018	PMID: 30042384	47	GSE89497	placenta	1f, 4d, Ext 11
	Pavlicev	2017	PMID: 28174237	48	GSE87726	placenta	1f
	Close	2017	PMID: 28279351	85	GSE93593	Neuro_diff	4d, Ext 11
	Guo	2015	PMID: 26046443	83	GSE63818	PGC	4d, Ext 11
	Velten	2017	PMID: 28319093	84	GSE75478	HSC	4d, Ext 11
	Enge	2017	PMID: 28965763	87	GSE81547	Pancreas	4d, Ext 11
	Darmanis	2015	PMID: 26060301	86	GSE67835	PFC	4d, Ext 11
Bulk RNA-seq	Zadora	2017	PMID: 28904069	54	available via approval	placenta	1f
	GTEx Consortium	2017	PMID: 29022597	88	phs000424.v6.p1	Human tissues	Ext 12
	D Antonio	2017	PMID: 28874753	55	E-MTAB-5714	293T	Ext 1a
	Chung	2018	PMID: 29395325	56	GSE99249	293T	Ext 1a
	Calderon	2019	PMID: 31570894	57	GSE118165	Immune cells	4d, Ext 11

ChIPseq	Barakat	2018	PMID: 30033119	59	GSE99631	hESC	1b
	Krendl	2017	PMID: 29078328	62	GSE105258	hESC	1c, Ext 1a-b
	Krendl	2017	PMID: 29078328	62	GSE105081	hESC	1c, Ext 1a- b
	Tsankov	2015	PMID: 25693565	58	GSE61475	hESC	1b, Ext 1d
	Dunn-Fletcher	2018	PMID: 30231016	61	GSE118289	placenta	1c
	Kwak	2019	PMID: 31294776	60	GSE127288	placenta	1c
DNAse-seq	Gao	2018	PMID: 29526463	63	GSA:CRA000297	embryo	Ext 1c
ATAC-seq	Wu	2018	PMID: 29720659	64	GSE101571	embryo	Ext 1c

Suppelemental table 2: Primer list

InFusion cloning primers		
ID	**sequence**	**description**
JF023	TTTTGGCAAAGAATTCATGGCCTGTATCTACCCA	pHCMV human suppressyn fwd
JF024	CCTGAGGAGTGAATTCTTATAGTTTTTGTATAAAGGAATGG	pHCMV human suppressyn rev
JF025	TTTTGGCAAAGAATTCATGGCCTGTATCTACCCAACC	pHCMV human suppressyn-HA fwd
JF026	CCTGAGGAGTGAATTCTTAAGCGTAATCTGGAACATCG	pHCMV human suppressyn-HA rev
JF046	TTTTGGCAAAGAATTCATGGCGCGTTCAACGCTCT	pHCMV amphoMLVenv fwd
JF047	CCTGAGGAGTGAATTCTCATGGCTCGTACTCTATGGGT	pHCMV amphoMLVenv rev
JF081	GCATAGTAGTCTCATTGCTACCACA	HERVH-1 (H62) provirus fwd
JF082	CATGACTCGGATCAGGGGAC	HERVH-1 (H62) provirus rev
JF091	ATCATTTTGGCAAAGaattCATGATCTTTGCTGGCAAGGCACC	pHCMV HERVH-1env fwd
JF093	CAGCCTGCACCTGAGGAGTGaattCctaagcgtaatctggaacatcgtatgggtaAGCTGAAGGGAGGTCTTGTGGTAAG	pHCMV HERVH-1env-HA rev
JF130	GGCGCCTAAGCTGGTATTCTTAACTATGTTGCTCC	pHIV7 eGFP downstream sequence fwd
JF131	CGCAGAGCCGGCAGCAGGCCGCGGGAAGGAAGGTCCGCTGGATTG	pHIV7 eGFP downstream sequence rev
JF135b	tttggggatcctgagcggccgcatccaaggtcggggcagga	U6 shRNA cloning fwd (pHIV-shSup cloning)
JF136	ccaaccactttctatactacacaaatatagaaaggttgggtagatacactttcgtcctttccacaagatataaagccaagaa	U6-shSup rev (pHIV-shSup cloning)

JF138	CTACTTCCTTTCGATACTACACAAATATCGAAAG GAAGTAGAAGTCCGGGCTTTCGTCCTTTCCACAA GATATATAAAGCCAAGAA	U6-shC rev (pHIV-shC cloning)
JF145	gtagtatagaaagtggttgggtagatacact ttttttgtaccgagctcggatccactagagatgg	shSup-CMV promoter fwd (pHIV7-shSup cloning)
JF147	GTAGTATCGAAAGGAAGTAGAAGTCCGGGCT ttttttgtaccgagctcggatccactagagatgg	shC-CMV promoter fwd (pHIV7-shC cloning)
JF152b	TGAACCGTCAGATCCGCTAGCATGGTGAGCAAGG GCGAGG	cerulean fwd (pHIV7-shRNA-cerulean cloning)
JF153b	AGAATACCAGttACTTGTACAGCTCGTCCATGCC	cerulean rev (pHIV7-shRNA-cerulean cloning)
JF154b	CTGTACAAGTAACTGGTATTCTTAACTATGTTGC TCC	cerulean-WPRE fwd (pHIV7-shRNA-cerulean cloning)
JF155	TGAACCGTCAGATCCGCTAGCATGGCCACCGAGT ACAAG	puroR fwd (pHIV7-shRNA-puroR cloning)
JF156	AGAATACCAGTTATTTGTAACCATTATAAGCTGC	puroR rev (pHIV7-shRNA-puroR cloning)
JF157	TTACAAATAACTGGTATTCTTAACTATGTTGCTC C	puroR-WPRE fwd (pHIV7-shRNA-puroR cloning)
JF165	cggtggccatGctagcggatctgacggttc	pHIV7-CMV promoter rev (pHIV-shRNA puroR cloning)
JF166	tgctcaccatgctagcggatctgacggttc	pHIV7-CMV promoter rev (pHIV-shRNA Cerulean cloning)
RAK010	TTTTGGCAAAGAATTCATGCTCTGCATCCTCATC CTCCT	pHCMV SMRVenv cloning primer fwd
RAK011	CCTGAGGAGTGAATTCCTACAGTCTGCCATATTC TAGGTCACGA	pHCMV SMRVenv cloning primer rev
qPCR primers		
JF042	GCAATTATTCCCCATGAACG	18S fwd
JF043	GGCCTCACTAAACCATCCAA	18S rev
JF108	CTGTGCATGCACATCGGTCACTG	Suppressyn fwd
JF109	GAGAAATTGGCCCAGACAAACACT	Suppressyn rev
JF112	ACCACGAACGGACATCCAAAG	Syncytin-1 fwd
JF113	GCCACTTTAACCGCAGTTGG	Syncytin-1 rev
Act-F	CGACAGGATGCAGAAGGAG	Actin fwd
Act-R	GTACTTGCGCTCAGGAGGAG	Actin rev
GUSB-F	AGAGTGGTGCTGAGGATTGG	GUSB fwd

GUSB-R	CCCTCATGCTCTAGCGTGTC	GUSB rev

1065

1066