A reversible haploid murine embryonic stem cell biobank resource for functional genomics

Ulrich Elling[1*,#], Reiner A. Wimmer[1*], Andreas Leibbrandt[1], , Thomas Burkard[1], Georg Michlits[1], Alexandra Leopoldi[1], Thomas Micheler[2], Dana Abdeen[1], Sergei Zhuk[1], Irene M. Aspalter[3], Cornelia Handl[1], Julia Liebergesell[1], Maria Hubmann[1], Anna-Maria Husa[1], Manuela Kinzer[1], Nicole Schuller[1], Ellen Wetzel[1], Nina van de Loo[1], Jorge Arturo Zepeda Martinez[1], David Estoppey[4], Ralph Riedl[4], Fengtang Yang[5], Beiyuan Fu[5], Thomas Dechat[6], Zoltán Ivics[7], Chukwuma A. Agu[1], Oliver Bell[1], Dieter Blaas[6], Holger Gerhardt[8-10], Dominic Hoepfner[4], Alexander Stark[11], and Josef M. Penninger[1,#]

[1] Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr Gasse 3, Vienna, Austria.
[2] Vienna Biocenter Core Facilities, Vienna Biocenter (VBC), Dr. Bohr Gasse 3, Vienna, Austria
[3] MRC Laboratory for Molecular Cell Biology and Institute for the Physics of Living Systems, University College London, UK.
[4] Novartis Institutes for BioMedical Research, Basel, Switzerland.
[5] Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
[6] Max F. Perutz Laboratories, Medical University of Vienna, Dr. Bohr Gasse 9, Vienna, Austria.
[7] Paul Ehrlich Institut, Paul Ehrlich Strasse 51-59, 63225 Langen, Germany.
[8] Max-Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany.
[9] German Center for Cardiovascular Research, Berlin, Germany.
[10] Berlin Institute of Health, Berlin, Germany.
[11] Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr. Bohr-Gasse 7, 1030 Vienna, Austria.

* The first two authors contributed equally to the work

[#]Correspondence and requests for materials should be addressed to U.E. (Ulrich.elling@imba.oeaw.ac.at) or J.M.P. (Email: josef.penninger@imba.oeaw.ac.at).

Summary

The ability to directly uncover the contributions of genes to a given phenotype is fundamental for biology research. However, ostensibly homogeneous cell populations exhibit large clonal variance[1,2] that can confound analyses and undermine reproducibility[3]. Here, we used genome-saturated mutagenesis to create a biobank of over 100,000 individual haploid murine embryonic stem cell (mESC) lines targeting 16,950 genes with genetically bar-coded, conditional and reversible mutations. This *Haplobank* is the largest resource of hemi-/homozygous mutant mESCs to date and is available to all researchers. Reversible mutagenesis overcomes clonal variance by permitting functional annotation of the genome directly in sister cells. We utilize Haplobank in reverse genetic screens to investigate the temporal resolution of essential genes in mESCs, and to identify novel genes that control sprouting angiogenesis and blood vessel lineage specification. Further, a genome-wide forward screen with Haplobank identified PLA2G16 as a host factor required for cytotoxicity by rhinoviruses, which cause the common cold. Thus, Haplobank clones and revertible technologies enable high-throughput, reproducible functional annotation of the genome.

Approaches to functionally analyze the mammalian genome include ENU mutagenesis[4], gene targeting[5], RNA interference[6,7,] and CRISPR-mediated genome editing[8]. Although powerful, these approaches suffer from various caveats, such as poor knockdown efficiency and off-target effects[9–11]. Additionally, clonal variability within populations can compromise comparisons and reproducibility[3,12,13,14]. Thus, reversibile mutations enabling direct comparison of phenotypes within a single clone are pivotal to study genetic dependencies.

To generate a conditional mutagenesis system at a genome-wide scale and at the clonal level, we applied insertional mutagenesis with genetically bar-coded lentiviral-, retroviral- [15,16] and transposon [*Tol2* [17] and *Sleeping beauty* (*SB*)[18]]-based vectors in haploid murine embryonic stem cells (mESCs, Fig. 1a), which enable recessive genetics. We analyzed several parthenogenic haploid mESC lines (129/Sv x C57/B6 background) (Extended Data Fig. 1a,b) and chose to pursue AN3-12 cells, which grow in feeder-free conditions and maintain a stable haploid genome across many generations in an undifferentiated state. These cells express pluripotency markers and differentiate into all germ layers *in vivo* (Extended Data Fig.1c-h). Importantly, AN3-12 cells display only minor genomic duplications and deletions, potentially affecting the genes *Cdh4*, *Taf4a*, *Agmo*, and *Cox7c* (Extended Data Fig. 2). Insertional mutagenesis permits the integration of invertable splice acceptors, resulting in conditional alleles, as well as high-throughput direct identification of integration sites. To map insertion sites and complex internal bar-codes (>10^7) by inverse PCR, an optimized universal sequencing strategy was established (Extended Data Fig. 3a,b). A combination of strategies was used to avoid genomic biases of the insertional mutagenesis systems (Fig. 1a,b; Extended Data Figs. 3c-f and 4a, reviewed in[19,20]), yielding genome-wide mutagenesis (Fig. 1c). Of note, *Tol2* outperformed the classical viral delivery systems and even *in silico* mutagenesis at <1 million integrations (Fig. 1c). The mutagenesis systems also generated anti-sense and intergenic integrations at high frequency due to cryptic transcriptional start sites (Extended Data Fig. 4b,c), however this is not expected to affect disruption of transcription in sense oriantation. Thus, we utilized various delivery systems to obtain unbiased, genome-saturated, and conditional mutagenesis.

Next, we developed a high-throughput pipeline to create a biobank of reversible mutations in AN3-12 haploid mESCs. Starting from haploid mESC pools that contained between 10^7-10^8 distinct mutations for each mutagen, we arrayed, processed, banked, and mapped over 100,000 individual mESC clones (Extended Data Fig. 4d-g). In total, we generated sense and anti-sense clones that target 16,970 of ~ 24,000 annotated mouse genes (genome release mm10), covering over 70% of the protein-coding genome (Supplementary Table 1 and Extended Data Fig. 5a-c). Integrations display a 5' bias in genes and in coding sequence (Extended Data Fig. 5d), resulting in truncations that are likely to generate loss-of-function alleles. All pools of mutated cells are available at www.haplobank.at. This resource represents a comprehensive library of mESC clones carrying hemi/homozygous, twice reversible, bar-coded integrations, combining the power of stem cells with tunable mutagenesis.

A key advantage of Haplobank is the ability to compare each mESC clone with its sister clone carrying the conditionally inverted splice acceptor. As a proof-of-principle, we analyzed two mESC clones containing sense integrations within *Ctnna1*, which encodes alphaE-catenin and is critical for mESC adhesion[21]. mESCs with sense integrations exhibited reduced *Ctnna1* expression and impaired cell adhesion, which were both restored by FlpO-mediated reversal of the integrated mutagenesis vectors. Further, Cre-mediated reversal back into the sense orientation again disrupted alphaE catenin-mediated cell adhesion (Fig. 1d,e). Additionally, we evaluated mESC lines carrying non-disruptive, anti-sense integrations in presumed essential genes. We infected these mESC lines with a pool of retroviruses that encode Cre and mCherry, or GFP only. If a gene is essential, Cre-mediated reversion of the integration to the disruptive sense orientation should specifically deplete $mCherry^+$ cells from the $mCherry^+/GFP^+$ cell pool over time, as detected by flow cytometry. Indeed, we confirmed the essential role of several genes for mESC survival (Fig. 2a, b). Thus, our system allows for a direct functional annotation of essentiality, instead of screening for the absence of mutations; moreover, one can directly examine the penetrance and timing of lethal phenotypes.

Next, we performed a genome-wide screen to uncover novel hits for resistance against infection with a common cold virus. We chose the rhinovirus serotype RV-A1a, which replicates in mouse cells[22]. A pool of mESCs carrying gene trap insertions was exposed to rhinovirus every other day for 3 weeks. The surviving, virus-resistant cells displayed an enrichment for multiple disruptive insertions in the low density lipoprotein receptor (*Ldlr*), a known entry portal for this virus (p-value=$2.9*10^{-12}$), and in the phospholipase *Pla2g16* (p-value=$1.4*10^{-11}$) (Fig. 2c). We confirmed that these genes are required for viral killing using three different sister clones with revertible integrations (Fig. 2d). Next, we used CRISPR-Cas9 technology to disrupt these genes in human embryonic kidney (HEK293T) cells and monitored competitive proliferation with and without RV-A1a infection. In this assay, LDLR was not required for RV-A1a-mediated killing (Fig. 2e), presumably because the virus can enter via other receptors in the absence of functional LDLR in HEK293T cells[23]. Importantly, we confirmed that inactivation of *PLA2G16* with different sgRNAs confers a selective survival advantage to HEK293T cells exposed to RV-A1a (Fig. 2e). Thus, Haplobank enables genome-saturated forward screening and validation to discover novel genes underlying specific phenotypes.

PLA2G16 can catalyze the rate limiting step of arachidonic acid synthesis, and thus couples to cyclooxygenases (COX) and prostaglandin synthesis[24,25]. However, different COX inhibitory drugs did not block RV-A1a-mediated cell death of control or repaired *Pla2g16* sister mESCs, nor did arachidonic acid enhance RV-A1a toxicity (Extended Data Fig. 6a). Selective survival of *Pla2g16* or *Ldlr* mutant mES cells was not affected by COX inhibitors (Extended Data Fig. 6b). Of note, the COX inhibitory drugs ibuprofen and indomethacin conferred partial resistance to RV-A1a in HEK293T cells (Extended Data Fig. 7a-c), albeit at concentrations that also affect other pathways[26]. Structurally, the short C-terminal vesicular domain of PLA2G16 extends into the endosomal lumen where the virion is located prior to releasing its RNA (Fig. 2f)[27]. To test the relevance of this domain, we edited the Pla2g16 coding region in the endosomal C-terminus and 3'UTR using CRISPR/Cas9 and selected for RV-A1a resistant cells. Mutations conferring resistance to

RV-A1a were enriched in the transmembrane domain and the vesicular domains (Fig. 2f; Extended Data Fig. 7d-f), consistent with a recent, independent haploid screen that identified PLA2G16 in picorna viridae infection[28]. Our results identify the C-terminal domain of Pla2g16 is a target to block rhinoviral infections.

As a third application of Haplobank, we investigated pathways required for angiogenesis[29]. Multiple candidate angiogenesis genes have been proposed, but few have been functionally validated[30,31]. We adapted embryoid body (EB) blood vessel sprouting, which recapitulates key features of *in vivo* angiogenesis[32], to our haploid mESCs. Sprouts stained positive for the endothelial marker CD31 and the basal membrane protein collagen IV, and formed lumens (Fig. 3b). Similar to *in vivo* blood vessel formation, the cells fronting the vascular structures exhibited characteristic features of tip cells, such as CD31-positive filopodia protrusions followed by stalk cells (Fig. 3b). Tip cells express Delta like ligand 4 (Dll4), which activates the Notch1 pathway on stalk cells to suppress their conversion into tip cells[33]. We used *Notch1* anti-sense non-disruptive clones from Haplobank and, via Cre recombination, created stable sister, sense clones that knock-out Notch1 expression (Extended Data Fig. 8a). EBs derived from multiple sense, knock-out sister clones displayed significantly increased vessel density (Extended Data Fig. 8b and not shown). Gene trap integration upstream of the Notch1 gene did not alter vessel density (Extended Data Fig. 8c). Thus, our EB sprouting assay recapitulates normal blood vessel development.

To identify novel genes, we selected candidates that are more highly expressed in tip versus stalk cells[30,31] and were associated with human vascular disease (Extended Data Fig. 8d). We focused on 32 genes,represented in Haplobank. To ensure that differences in angiogenesis are directly linked to inactivation of the respective target genes, and not genetic background noise or clonal effects, we performed color tracing. We infected selected mESCs with GFP- or mCherry-Cre expressing retroviruses to generate sister clones with reverted orientation of the integration (Fig. 3c). We observed large variability between independent clones (Extended Data Fig. 8e,f), highlighting the importance of comparing mutant to genetically repaired sister clones for each gene. Genetic inactivation

of *Myst3, Mecom, Gja1, Gabrb3, Tnfrsf1a,* and *Dlg2* reduced sprouting angiogenesis and decreased vessel formation at least 2-fold compared to anti-sense clones; gene trapping of *Enpp3, Smarca1, Ndufs4, Plcb1,* or *Hck* promoted blood vessel growth (Fig. 3d,e; Extended Data Fig. 9). Thus, Haplobank enabled rapid, functional and reproducible validation of candidate angiogenesis genes in engineered blood vessels.

To assess the *in vivo* role of these novel angiogenesis genes, we generated EBs from mESCs stably expressing mCherry and injected them into immuno-compromised mice. The resulting teratomas were assessed for mESC-derived mCherry$^+$ IB4$^+$ blood vessels (Fig. 4a, Extended Data Fig. 10a). To control for teratoma growth rates, we injected 1:1 mixed mosaic EBs of sense (mutant, GFP+) and their respective anti-sense clones (repaired, mCherry+). The contribution of GFP- or mCherry-expressing cells to the endothelial lineage as well as non-endothelial tissues was assessed by cytometry (Extended Data Fig. 10b). Mutated clones of *Myst3*, *Gja1* and *Grin2b* displayed a decreased contribution to IB4$^+$ vasculature (Fig. 4a; Extended Data Fig. 10b). In contrast, mutant clones that showed increased vessel density *in vitro* exhibited a greater capacity to form blood vessels *in vivo* than their anti-sense sister clones (Fig. 4a). These data were confirmed using *in situ* blood vessel analysis of mixed teratomas (Fig. 4b). To test whether the identified genes modulate angiogenesis via specification of tip cell fate, we performed mosaic tip cell competition assays (Extended Data Fig. 10c,d). As a positive control, we assessed genetic modulation of the Notch1 pathway. As expected, mCherry$^+$ (sense knock-out) *Notch1* mutant cells preferentially localized to the tip position compared to GFP$^+$ anti-sense Notch1-expressing sister clones (Fig. 4c,d). Most clones carrying gene mutations that increased sprouting activity displayed a significant increase to the tip cell position, and *vice versa* (Fig. 4c-e). Therefore, the newly identified angiogenesis genes can control the tip cell fate.

We tested whether one of our novel angiogenesis genes, *Gja1,* encoding the gap junction protein Connexin43 [34], is involved in physiological vascularization of the mouse retina, which begins at birth and progresses until postnatal day 7 [35]. At postnatal day 6, we observed high Gja1/Connexin43 expression at the angiogenic front where

Gja1/Connexin43 localized to endothelial junctions, with the highest intensity in tip cells (Extended Data Fig. 10e). At the vascular plexus, Gja1/Connexin43 expression was predominantly detected in perivascular cells, not endothelial cells (Extended Data Fig. 10f). Moreover, Gja1/Connexin43 was primarily detected at the tip of the developing vascular sprout in 3D blood vessels (Extended Data Fig. 10g). Newborn mice injected i.v. with a Gja1/Connexin43 blocking peptide displayed a delay in vascular network progression and complexity in the retina compared to those injected with a scrambled control peptide[36] (Fig. 4f). The number of tip cells at the angiogenic front was decreased, together with decreased numbers of branch points in the vascular plexus (Fig. 4f). Thus, Gja1/Connexin43 is a key regulator of tip cell fate and physiological angiogenesis *in vivo*.

In summary, the Haplobank resource contains over 100,000 individually mutagenized and bar-coded mESC lines targeting 16,970 protein coding genes. Haplobank complements a collection of 3,396 reversibley targeted genes in a near haploid human leukemia cell line[45]. Our proof-of-principle experiments uncovering genes required for rhinovirus infection and angiogenesis establish the power of Haplobank in forward and reverse genetic screens, respectively. The strong variability between independent clones revealed the importance of assessing mutant and repaired clones side-by-side, and addresses an increased demand for rigor and reproductibility[37]. Thus, Haplobank clones and revertible technologies enable high-throughput, reproducible functional annotation of the genome.

Supplementary information is available in the online version of the paper.

Acknowledgments. We thank all members of our laboratories, IMBA/IMP and VBCF services for support and Life Science Editors for assistance. We thank Britta Knapp, Ireos Filipuzzi, and Thomas Aust for clone picking, N. Rao Movva and Tewis Bouwmeester (NIBR) for support, and Kristina Handler for the differentiation protocols. Haplobank is generously funded by the Austrian National Bank (OeNB), an Advanced ERC grant and Era of Hope/National Coalition against Breast Cancer/DoD (to J.M.P.). U.E. is a Wittgenstein Prize fellow. D.B is supported by FWF P23308-B13. A.S. is supported by an ERC Consolidator Grant , Boehringer Ingelheim, and FFG.

Author Contributions.

U.E. generated the haploid library with technical support from An.L., C.H., J.L., M.H., A.-M.H., M.K., N.S., E.W., N.V.D.L, D.H., R.R., D.E.. U.E., R.AW and Al.L. characterized cell line. An.L., G.M., U.E., D.B., T.D., performed Rhinovirus work. A.S., T.B., and T.M wrote bioinformatics algorithms and set-up the Haplobank webpage. S.Z. performed RACE experiments, F.Y. and B.F. karyotyping, C.A.A. supported standardization. J.A.Z.M. and O.B. performed ATACseq. Z.I. advised on mutagenesis vectors. R.A.W., I.A., D.A., Al.L, and H.G. performed blood vessel experiments. U.E. and J.M.P. coordinated the project.

Competing financial interests

The authors declare no competing financial interests.

References

1. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14,** 357–68 (2013).
2. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26,**

304–19 (2016).
3. Begley, C. G. & Ellis, L. M. Drug development: Raise standards for preclinical cancer research. *Nature* **483,** 531–3 (2012).
4. Justice, M. J., Noveroske, J. K., Weber, J. S., Zheng, B. & Bradley, A. Mouse ENU mutagenesis. *Human Molecular Genetics* **8,** 1955–1963 (1999).
5. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323,** 445–8 (1986).
6. Elbashir, S. M. *et al.* Duplexes of 21 ± nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494–498 (2001).
7. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296,** 550–553 (2002).
8. Jinek, M. *et al.* A Programmable Dual-RNA – Guided. *Science* **337,** 816–822 (2012).
9. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–6 (2013).
10. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR / Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34,** 1–4 (2016).
11. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34,** 11–14 (2016).
12. Rouhani, F. *et al.* Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. *PLoS Genet.* **10,** (2014).
13. Fedorov, L. M., Haegel-Kronenberger, H. & Hirchenhain, J. A comparison of the germline potential of differently aged ES cell lines and their transfected descendants. *Transgenic Res.* **6,** 223–231 (1997).
14. Echeverri, C. J. *et al.* Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat. Methods* **3,** 777–9 (2006).
15. Schnutgen, F. *et al.* Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* **102,** 7221–7226 (2005).
16. Schnütgen, F. *et al.* Enhanced gene trapping in mouse embryonic stem cells. *Nucleic Acids Res.* **36,** (2008).
17. Ika Mayasari, N. *et al.* Mixture of differentially tagged Tol2 transposons accelerates conditional disruption of a broad spectrum of genes in mouse embryonic stem cells. *Nucleic Acids Res.* **40,** (2012).
18. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvák, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* **91,** 501–510 (1997).
19. Bellen, H. J. *et al.* The Drosophila gene disruption project: Progress using transposons with distinctive site specificities. *Genetics* **188,** 731–743 (2011).
20. Ivics, Z. *et al.* Transposon-mediated genome manipulation in vertebrates. *Nat. Methods* **6,** 415–422 (2009).
21. Torres, M. *et al.* An alpha-E-catenin gene trap mutation defines its function in preimplantation development. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 901–6 (1997).
22. Reithmayer, M., Reischl, A., Snyers, L. & Blaas, D. Species-specific receptor

recognition by a minor-group human rhinovirus (HRV): HRV serotype 1A distinguishes between the murine and the human low-density lipoprotein receptor. *J. Virol.* **76,** 6957–6965 (2002).
23. Hofer, F. *et al.* Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc. Natl. Acad. Sci.* **91,** 1839–1842 (1994).
24. Jaworski, K. *et al.* AdPLA ablation increases lipolysis and prevents obesity induced by high-fat feeding or leptin deficiency. *Nat. Med.* **15,** 159–168 (2009).
25. Duncan, R. E., Sarkadi-Nagy, E., Jaworski, K., Ahmadian, M. & Hei, S. S. Identification and functional characterization of adipose-specific phospholipase A2 (AdPLA). *J. Biol. Chem.* **283,** 25428–25436 (2008).
26. Tsutsumi, S. *et al.* Endoplasmic reticulum stress response is involved in nonsteroidal anti-inflammatory drug-induced apoptosis. *Cell Death Differ.* **11,** 1009–1016 (2004).
27. Uyama, T. *et al.* Interaction of Phospholipase A/acyltransferase-3 with Pex19p: A possible involvement in the down-regulation of peroxisomes. *J. Biol. Chem.* **290,** 17520–17534 (2015).
28. Staring, J. *et al.* PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* **541,** 412–416 (2017).
29. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146,** 873–887 (2011).
30. Del Toro, R. *et al.* Identification and functional analysis of endothelial tip cell-enriched genes. *Blood* **116,** 4025–4033 (2010).
31. Strasser, G. A., Kaminker, J. S. & Tessier-Lavigne, M. Microarray analysis of retinal endothelial tip cells identifies CXCR4 as a mediator of tip cell morphology and branching. *Blood* **115,** 5102–5110 (2010).
32. Jakobsson, L. *et al.* Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat. Cell Biol.* **12,** 943–953 (2010).
33. Hellström, M. *et al.* Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature* **445,** 776–780 (2007).
34. Valdimarsson, G., De Sousa, P. A., Beyer, E. C., Paul, D. L. & Kidder, G. M. Zygotic expression of the connexin43 gene supplies subunits for gap junction assembly during mouse preimplantation development. *Mol. Reprod. Dev.* **30,** 18–26 (1991).
35. Pitulescu, M. E., Schmidt, I., Benedito, R. & Adams, R. H. Inducible gene targeting in the neonatal vasculature and analysis of retinal angiogenesis in mice. *Nat. Protoc.* **5,** 1518–34 (2010).
36. Chaytor, A. T., Martin, P. E., Edwards, D. H. & Griffith, T. M. Gap junctional communication underpins EDHF-type relaxations evoked by ACh in the rat hepatic artery. *Am. J. Physiol. Heart Circ. Physiol.* **280,** H2441–H2450 (2001).
37. Baker, M. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* **533,** 452–454 (2016).

Figure legends

Figure 1. A repairable mutant mES cell library.

a) Schematic representation of insertional mutagenesis vectors: Splice acceptor sites (SA) are revertible using non-compatible loxP/lox5171 and FRT/F3 sites (triangles). G418 resistance is conferred by beta-Geo (bgeo) transcribed from the revertible cassette (gene trap vectors, GT) or Neo independently from a PGK promoter (polyA trap, pA), stabilized by a splice donor (SD). Six osteopontin enhancer (OPE) elements (enhanced gene trap; Lenti-ETG, Retro-EGT, and *Tol2*-EGT vectors) enhance expression of beta-Geo via Oct4/Pouf51 binding. RetroRS carries a spacer sequence between loxP sites and lacks OPEs. Purple diamonds indicate internal barcodes (BC). LTR, long terminal repeats; L200/R175 and LITR/RITR, terminal repeats of *Tol2* and *SB*. **b)** Heat map representing numbers of integrations per gene per 1 million integrations. Gene expression levels are shown (blue=highly expressed, white=not expressed). Color code shows numbers of integrations. **c)** Saturation of mutagenesis systems compared to random *in silico* mutagenesis. Y-axis, total numbers of insertions versus % of genes with integrations. **d)** Schematic representation of splice acceptor inversions. **e)** Loss of mESC adhesion in clones with integrations in intron 1 of *Ctnna1*. Inversion of the gene trap restores cell adhesion, subsequent reversion again disruptes adhesion. Phalloidin images polymerized actin; DAPI visualizes nuclei. Size bars, 10 μm. One representative experiment out of 2 biological replicates is shown.

Figure 2. Essential genes for mESC and common cold virus infections.

a,b) Functional annotation of essential mESC genes. **a)** Competitive growth assays of anti-sense (GFP^+) and Cre-reverted sense ($mCherry^+$) sister cells harboring integrations in the indicated genes. Cell populations were analyzed at the indicated days after Cre addition using flow cytometry. Means +/- SD of biological triplicates. **b)** FACS plots for the essential gene *Psmd1* illustrating depletion of $Cherry^+$ cells. **c)** Integration sites of top scoring genes in our haploid mESC survival screen of human rhinovirus RV-A1a infections. Loss of function score for integrations into the *Ldlr* locus $p=2.9x10^{12}$ and $p=1.4x10^{11}$ for *Pla2g16*. Sense integrations, red triangles; anti-sense integrations, green; exons, blue boxes. Transcriptional start sites are marked. **d)** Growth advantage of sense versus respective anti-sense sister mESC harboring integrations in *Pla2g16* or *Ldlr* upon infection with RV-A1a. In un-infected cells, mutation of these genes did not confer growth advantages; arbitrarily set to 1. **e)** Human embryonic kidney HEK293T cells were transduced with 4 different sgRNAs against *PLA2G16* and *LDLR* in biological triplicates, mixed with control GFP^+ HEK293T cells at a ratio of 1:3. Ratios of control to mutated HEK293T cells were evaluated on day 13 after infection using FACS. Data in d and e are means +/- SD, normalized to uninfected cells. Individual data points and error bar (STDEV); one tailed students t-test, *< 0.1, **< 0.01, ***< 0.001. **f)** Targeting of the C-terminal Pla2g16 domain using CRISPR/Cas9. Upon selection of haploid cells to ensure hemizygous editing, cells were split and maintained in the presence and absence of RV-A1a.

Figure 3. Novel regulators of angiogenesis.

a, Generation of sprouting vasculature from haploid mESCs, differentiated into Embryoid Bodies (EBs) and cultured with VEGF-A (30ng/mL). **b**, CD31$^+$ (green) endothelial cells and filopodia, indicative of tip cells in blood vessel sprouts. Luminal structures and Collagen IV$^+$ basement membranes are shown in right panels. Scale bars, 200μm or as indicated. **c,** Schematic outline for functional validation of candidate genes in sprouting angiogenesis. Haplobank clones were infected with GFP or mCherry-Cre viruses to generate disruptive sense (S) and anti-sense (AS) sister cells. **d,** Representative images of hypo- and hyper-sprouting sense (S) and anti-sense (AS) sister clones. IB4 marks endothelial cells. Scales, 500μm. **e,** Quantification of IB4$^+$ blood vessel sprouts. Data were normalized to the respective anti-sense sister clones. Means ± SEM are shown from a minimum of n=3 independent experiments. *<0.05; **< 0.01; ***< 0.001 (Two tailed students t-test).

Figure 4. *In vivo* angiogenesis and cell specifications.

a, Quantification of the indicated sense and anti-sense targeted clones to form blood vessels in teratomas. Means ± SEM are shown (n=3 independent teratomas; n=5 for the AN3-12 control). *< 0.05 (Two tailed students t-test). **b,** *Gja1*-sense (GFP$^+$) and *Gja1*-antisense (mCherry$^+$) sister mES cells were mixed 1:1 to form chimeric EBs and subsequently injected into immunocompromised *nu/nu* mice. Representative sections to identify IB4, GFP, and mCherry expressing cells are shown. Scale bar 50μm. **c,** *Notch1* anti-sense and mCherry-Cre$^+$ *Notch1*-sense sister mES cells were mixed (1:1 ratio) to generate mosaic blood vessels and analyzed for red or green cells at the tip position. Scale bar 200μm. **d,** Relative tip cell position of sister cells with sense and anti-sense integrations

in the indicated genes, determined in chimeric 3D sprouts. Means ± SEM of a minimum of n=3 independent experiments. *< 0.05; **< 0.01; ***< 0.001 (Two tailed students t-test). **e,** Representative image of mosaic blood vessel sprouts from *Gja1* sense and anti-sense sister clones. Scale bars, 500 μm and 100 μm (insert). **f,** Intravenous injection of a Gja1 inhibitory peptide (GAP26) into neonatal mice abrogated retinal angiogenesis. At day 5 after birth, retinas were isolated and stained for $IB4^{+}$ blood vessels. Scale bars: 500μm upper, 100μm lower panels.

Supplementary Information

Methods

Haploid ES cell cultures. The murine haploid ES cells used for this study are a feeder independent clonal derivative of HMSc2 termed AN3-12. Haploid ES cells were grown in standard ES cell medium with serum and LIF. Cell culture-grade dishes were from Greiner (15 cm dishes, 15cm Cellstar cell culture dishes, cat no 639160) and NUNC (all other formats, e.g. 10cm dish Nunclon Δ Surface, cat no. 150350; 6-well Nunclon Δ Surface, cat no. 140675). The following ES cell medium (ESCM) was used: 450 ml DMEM (Sigma D1152); 75 ml FCS (Invitrogen); 5.5 ml P/S (Sigma P0781); 5.5 ml NEAA (Sigma M7145); 5.5 ml LGlu (Sigma G7513); 5.5 ml NaPyr (Sigma S8636); 0.55 ml βME (Merck 805740; dilute 10 µl bME in 2.85 ml PBS for a 1000x stock); and ESGRO (Millipore ESG1107; used according to instructions by the manufacturer). For freezing, ES cell clones were expanded, harvested from a confluent 10 cm dish by trypsinization, the reaction stopped with ESC medium, and subsequently the cells centrifuged at 1200 rpm for 5 min. Supernatants were discarded and cell pellets re-suspended in 1 ml PBS. Twenty µl of the cell suspension were removed and used to prepare DNA for barcode and integration site PCR. Finally, 7 ml freezing medium was added (1 vol ESC medium, 1 vol FCS, 11% DMSO (Sigma # 41648)) and cells were frozen in quadruplicates in 2ml cryovials. Enzymatic Mycoplasma tests were performed on a weekly basis. For purification and maintenance, haploid cells were trypsinized and incubated in 10 µg/ml Hoechst33342 (Sigma B2261) for 30-40min at 37˚C and subjected to FACS sorting using a BD AriaIII equipped with a near UV laser. For flow cytometric analysis cells were trypsinized and

analyzed using a FACS BD LSRFortessa (BDBiosciences) equipped with a high throughput sampler. Data analysis was performed using FACS Diva and FlowJo. To generate single cell-derived colonies, 200 cells/10cm dish were plated and grown for 10-11 days. Cells were washed once with PBS, and clones manually picked in 20 µl PBS under a binocular into 96U well plates (Nunc). The picked colonies were dispersed into single cells by incubation with 5 µl 0.25% trypsin in 20 µl PBS for 6 min at 37°C. The reaction was stopped with 175 µl ESCM, and the cells were then split into 96F well plates, or directly into a 24 well dish for further analysis. For robot assisted picking, colonies were transferred in 20µl PBS or DMEM in U-well plates subsequent to picking into a Hamilton Lab Star. Cells were dissociated using trypsin as above and plated in 5 replicas using the 96 head pipetting block. Clones were expanded in 96 well plates for 3 days and frozen in matrix plates with 2D barcodes or in cryogenic vials (Thermo Scientific) and finally transferred to liquid nitrogen. A step-by step protocol describing haploid cell cultivation can be found at Protocol Exchange[38]

Viral and transposon vectors for haploid ES cell mutagenesis. For retro- and lentiviral library generation, gene trap viruses carrying a neomycin resistance were packaged in PlatinumE (Cell Biolabs, used only at early passage after obtaining from vendor, tested for cell identity by antibiotics resistance and capability to package virus) or LentiX (Clontech, used only at early passage after obtaining from vendor, further identified by capability to package virus) cells, respectively. For some experiments, virus was concentrated by centrifugation (25.000 rpm, 4°C, 4 hrs) and FACS sorted haploid ES cells were then infected for 8 hrs in the presence of 2µg polybrene per ml; 50 µM JQ1 was used to increase infection efficiency. For transposon based library generation, FACS-sorted haploid ES cells were

nucleofected (Amaxa Nucleofector, program A13) using transposons containing gene trap cassettes carrying a splice acceptor, a RPB1 promoter and a neomycin resistance gene. Thirty hrs post infection, selection for gene trap insertions started using G418 (Gibco) at 0.2mg/ml. To estimate numbers of integrations and thus library complexity, 10,000 cells (for viruses) or 50,000 cells (for transposons) were plated on 15 cm dishes and selected using G418. For comparison, 1000 cells were plated but not exposed to G418 selection. On day 10, colonies were counted.

4D pooling. In brief, up to hundred 96-well plates containing lysed cells were stacked into 10 towers of 10 plates (i.e. slices) height. In the first step, lysates were pooled into "Master tower" plates containing pools of all plates within 1 tower and a fraction of lysate was separated into jars containing pools of all wells of each plate termed "slice". While processing further towers, identical slices were united with previous ones and new master tower plates were generated. In a second run, master tower plates were pooled along *rows* and *columns* of 96 well plates representing all wells within 1 tower. Finally, this resulted in 40 pools representative of 12 columns, 8 rows, 10 slices, and 10 towers. Lysate of each well were thus present in 4 coordinates. Upon DNA purification, the 40 individual pools were processed for inverse PCR adding individual indices to each pool in the PCR step. Subsequently, the pool of all coordinates was sequenced within one lane of an illumina sequencer. Two independent strategies of iPCR based on different restriction enzymes were used to identify mutations and coordinates. In addition, direct genomic PCR amplifying the internal barcodes of the mutagens was performed with the same experimental indices. For details see below.

Parallel cloning of internal barcodes. All mutagenesis vectors were designed with highly complex barcodes (BC) for confirmation of the generated clones. The complex BC sequences were obtained as ultramers from 'Integrated DNA Technologies' as "PAGE Ultramer DNA Oligos" and contained 32 bp BC sequences flanked by BspEI and BstEII restriction sites. Two versions with either strong-weak (SW) or weak-strong (WS) alternating sequences were cloned into each frame; BC sequences were amplified by PCR and purified over columns (Qiagen). The gene trap vector plasmids (GT-MCS, frame 0/1/2) were digested with AgeI and BstEII and dephosphorylated. Fragments were separated by agarose gel electrophoresis, phenol extracted and precipitated with ethanol. Amplified BC-oligos were digested with BspEI and BstEII. Backbones and inserts were ligated with T4 DNA ligase (NEB) and electroporated (BioRad, Gene Pulser II) into appropriate electrocompetent cells. Bacteria were plated and grown over night and further expanded in liquid culture for 6h. Plasmid DNA extraction was performed using established protocols (Qiagen). The following barcode PCR primers were used: forward primer: GTTGATCTGAGCTACTCATCAACGGT; reverse: AAGTTCCTTCTGGTTCTGGCTCTGCT. Twenty µl of PCR reactions were analyzed on a 2 % agarose gel and PCR products were purified for sequencing using Illustra ExoStar 1 – step kit (GE Healthcare). For Sanger sequencing, the barcode PCR-reverse primer was used.

Mapping of genomic integrations by next generation sequencing. Enzyme 1 (E1) was used to fragment the genome. As the recognition sequence for E1 is also present adjacent to the barcode of the gene trap vector, it is possible to retrieve the exact integration site of the gene trap cassette within the genome by circularizing E1-digested gDNA and subsequent amplification of the genomic region by inverse PCR (iPCR) using primers "US" and "DS". To

improve iPCR efficiency, a linearization step using E2 was introduced, which re-opens the rings generated previously. Two mapping strategies using enzymes E1 were implemented for each mutagenesis system. For protocol details and sequence information, please also see www.haplobank.at.

Mutagenesis system	Enzyme1 (E1)	Enzyme2 (E2)	Mapping Strategy
Retroviruses	NlaIII, MseI	SbfI	5'
Tol2	NlaIII, TaqI	PacI	3'
Lentiviruses	NlaIII, TaqI	PacI	5'

The following illumina iPCR primer sequences were used: DS: AATGATACGGCGACCACCGAGATCTACACGAGCCAGAACCAGAAGGAACTTGAC
US: CAAGCAGAAGACGGCATACGAGATINDEXGTGACTGGAGTTCAGACGTGTGCTCTTC
INDEX = custom barcode, 4-8bps.

For genomic DNA preparation, cell pellets were lysed in lysis buffer (10mM Tris-HCl pH 8.0, 5mM EDTA, 100mM NaCl, 1% SDS, 1 mg/ml Proteinase K) and incubated at 55˚C overnight. RNAse A (Qiagen, 100mg/ml) was added at a ratio of 1:1000. After incubation for 1 hr at 37˚C two phenol/chloroform/isoamyl alcohol extractions and a chloroform/isoamyl alcohol extraction, using phase lock tubes (5Prime), were performed. Samples were then incubated over night at the digestion enzyme specific temperatures. Restriction digests were purified using a PCR Purification Kit (Qiagen) and Ring ligated using T4 DNA ligase (Roche Applied Science). The ring-ligation reaction was incubated over night at 16˚C, heat-inactivated at 65˚C for 15 min and linearized by adding 1 µl Enzyme2 for 2 hrs at 37˚C. The digest was again purified using a PCR Purification Kit (Qiagen). The eluate was then used for iPCR. Twenty µl sample were finally analyzed on an agarose gel and the remaining PCR products purified using a Gel Extraction Kit (Qiagen). Next generation sequencing (NGS) was performed on an

illumina HiSeq2500 sequencer according to the manufacturers protocols. Sequencing primers used for the 1st read as well as the experimental indices in BC-PCR were custom-made, other primers were standard illumina primers:

- Retrovirus 1:1 mix (due to alternative processing of viral LTRs):

  GAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA

  TGAGTGATTGACTACCCACGACGGGGGTCTTTCA
- Tol2: CACTTGAGTAAAATTTTTGAGTACTTTTTTACACCTCTG
- Lentivirus: CAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA

To additionally confirm the genomic integration site of the gene trap cassettes, gene specific primers flanking the mapped locus were paired with mutagen specific primers to confirm integration as well as absence of the wild type allele. Integration site-specific primers were designed using Clone Manager 9 software (Sci-Ed Software). For each integration site, 4 PCR reactions with respective primer combinations (integration site-specific forward and reverse primer and integration site specific forward or reverse with mutagen-specific forward or reverse primer) were performed. The following mutagen specific primers were used:

Retrovirus-INT-fwd: CCAGAACCAGAAGGAACTTGC

Retrovirus-INT-rev: TACAGACGCAGGCGCATAACAC

Lentivirus-INT-fwd: GCCAGAACCAGAAGGAACTTGC

Lentivirus-INT-rev: AGAGCTCCTCTGGTTTCCCTTTC

Tol2-INT-fwd: GAGCCAGAACCAGAAGGAACTG

Tol2-INT-rev: CCGGGCAATGGATTGATATTGG

4D deconvolution of NGS reads. Based on the restriction enzyme used for genome digestion, the 1st read was cut at the first restriction site if present within the 50bp of mapped sequence. A minimal length of 15nt was required. The sequence was aligned with bowtie against the mm10 genome[39]. A maximum of 1 mismatch (-v 1) was allowed. Only unique alignments (-m 1) in the best strata are reported. The 2nd read reading out the internal barcode was trimmed after 32nt. The alignment/internal barcode pairs were assigned to a defined well via the experimental indices. A valid assignment required a minimal count of 3 per coordinate and a minimal fraction of 0.75 over alternative coordinates. The internal barcodes were merged with 4D pooling data if its count was higher than 20. This strategy was also used to verify the existing coordinates and uncover missing coordinates and/or additional integrations. Mapped insertions were intersected with the annotation (Ensembl) of all gene features including the intergenic region using BEDtools[40]. *In silico* insertions were randomly sampled from the mouse genome and aligned back to the genome.

Chromosome spreads and FISH. Metaphase chromosomes were collected according to standard protocols. For multiplex-fluorescence in situ hybridization (M-FISH), Mouse 21-color painting probe was made following the pooling strategy [2]). The M-FISH probe was denatured at 65°C for 10 min before being applied onto the denatured slides. The hybridization area was sealed with a 22 x 22-mm coverslip and rubber cement. Hybridization was carried out in a 37°C incubator for 2 nights. The post hybridization washes included a 5-min stringent wash in 0.5x SSC at 75°C, followed by a 5-min rinse in 2 x SSC containing 0.05% Tween20 (VWR) and a 2-min rinse in 1x PBS, both at room temperature. Finally, slides were mounted with SlowFade Gold mounting solution containing 4'6-diamidino-2-phenylindole

(DAPI, Invitrogen). Images were visualized on a Zeiss AxioImager D1 fluorescent microscope equipped with narrow band-pass filters for DAPI, DEAC, FITC, CY3, TEXAS RED, and CY5 fluorescence and an ORCA-EA CCD camera (Hamamatsu). M-FISH digital images were captured using the SmartCapture software (Digital Scientific UK) and processed using the SmartType Karyotyper software (Digital Scientific UK). Approximately 200 metaphases from AN3-12 were counted for the diploid *vs.* haploid frequency and 10 well-spread metaphases were fully karyotyped by M-FISH and DAPI-banding pattern[41].

CNV analysis. Discriminative coverage analysis was performed to identify differences between genotypes in the number of copies of a genomic region as previously described (PMID:22136931). In brief, reads from 129, AN312 and C57BL/6J samples were aligned against the mm10 C57BL/6J mouse reference genome using bowtie v1.2 with parameters "-v 3 -m 1 --best --strata". Coverage in non-overlapping 10k windows over the whole genome was calculated using bedtools (v2.26.0) makewindows and multicov (PMID:20110278). Read counts were scaled by library size and for each sample windows with extreme outlier counts were removed.

SNP mapping. Variants were called with GATK v3.7-0 (PMID: 20644199) following established best practices. In brief, reads were aligned to the mm10 C57BL/6J mouse reference genome using bwa mem (bwa v0.7.12), duplicate reads were marked using Picard v2.6.0, reads were locally realigned around indels, base quality recalibration was performed using mm10 dbSNP as a training set, variants were called using HaplotypeCaller with ploidy set to 1 for AN312 and 2 for C57BL/6J and 129 samples, and variant filtration was applied

according to GATK guidelines. For each sample, homozygous SNPs were counted in nonoverlapping 10-kb sliding windows and counts were plotted to estimate the distribution SNP densities on the mm10 reference genome.

RNAseq. Reads were screened for ribosomal RNA by aligning with BWA (v0.6.1)[42] against known rRNA sequences (RefSeq). The rRNA subtracted reads were aligned with TopHat (v1.4.1)[43] against the *Mus musculus* genome (mm10); a maximum of 6 mismatches were allowed. Maximum multi-hits were set to 1 and InDels as well as Microexon-search was enabled. Additionally, a gene model was provided as GTF (UCSC RefSeq mm10). rRNA loci were masked on the genome for downstream analysis. Aligned reads were subjected to FPKM measurements with Cufflinks (v1.3.0)[44,45]. Furthermore, only those fragments compatible with UCSC RefSeq annotation (mm10) of genes with at least one protein coding transcript were allowed and counted towards the number of mapped hits used in the FPKM denominator. The mean FPKM was calculated over the replicates.

Chromatin-IP. ChIP was essentially performed as described previously[46]. Briefly, after trypsinization, 25 million haploid ES cells were washed once in PBS before fixation for 7 min at room temperature by addition of formaldehyde to a final concentration of 1%. Crosslinking was quenched by addition of 2.5 M glycine (0.125 M final concentration) and cells were then incubated on ice. Crosslinked cells were spun at 600 g for 5 min and nuclei prepared by consecutive washes with Rinse 1 buffer (final: 50 mM HEPES pH 8.0, 140 mM NaCl, 10% glycerol, 0.5% NP40, 0.25% Triton X100, 1 mM EDTA) followed by Rinse 2 buffer (final: 10 mM Tris pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Nuclei were re-suspended in 2 ml

total volume of sonication buffer (0.1% SDS, 1mM EDTA, pH 8, 10mM Tris HCl, pH 8 with protease inhibitors complete mini (Roche)) and then sonicated with a Covaris E22 sonicator (Covaris settings: 5 % duty cycle, PIP 140, 200cyles/ burst, 15 min). ChIP was performed at 4 degrees in ChIP buffer (final: 50 mM HEPES/KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 % Triton X100, 0.1 % DOC, 0.1% SDS) with the indicated antibody. Precipitated chromatin was purified using Dynabeads Protein G (Thermofisher, 10003D).

ATACseq. ATACseq was performed as reported previously[47]. Briefly, 50,000 cells were washed on cold PBS before gentle resuspension in 50 µl of cold lysis buffer (10 Mm Tris-HCl pH 7.4, 10mM NaCl, 3 mM MgCl2, 0.1% Igepal CA-630). For transposition reaction, nuclei were collected by immediate centrifugation for 15 min at 4^{o}C. Supernatant was removed and nuclei were suspended in 50 µl of transposition reaction mixture (25 µl 2x TD buffer (Nextera kit), 2.5 µL TDE1 (Nextera Tn5 Transposes), 22.5 µL nuclease-free water). Transposition reaction was performed at 37^{o}C for 30 minutes followed by DNA purification using Qiagen MinElute Kit.

5'RACE. 10 µg of total RNA was isolated from ES cells using Trizol (Thermo Fisher Scientific), followed by additional purification and DNAse treatment with QIAGEN RNeasy mini kit. Poly-A mRNA was enriched using dT(25)-Dynabeads (Thermo Fisher Scientific) using manufacturers recommendations. Purified mRNA was reverse transcribed with Super Script II polymerase (Invitrogen) in the presence of random hexamer primers using the following incubation protocol: 25^{0}C – 10min, 42^{0}C – 50min, 50^{0}C – 10min, 70^{0}C – 15min. cDNA from the reaction was purified with MinElute reaction clean-up kit (QIAGEN) and tailed with

dCTPs using Terminal Transferase (NEB). 1/10 of the reaction was used for the first round of PCR with dG-anchor-primer/g.s.-outer-primer: 96^{0}C – 2min, (96^{0}C -15s; 60^{0}C – 15s; 68^{0}C – 3min)x23, 72^{0}C – 10min. Reaction products were diluted 10 X and used for the second round of PCR with dG-sec-anchor /g.s.-inner-primera with following incubation protocol: 96^{0}C – 2min, (96^{0}C -15s; 63^{0}C – 15s; 72^{0}C – 3min)x23, 72^{0}C – 10min. Reaction products were purified with MinElute clean-up kit and cloned into a pJet1.2 vector (Thermo Fisher Scientific). Bacterial colonies were picked and sequenced with beta-geo_R4 primer. The following primers were used:

g.s.-inner-pr	737	GTATCGGCCTCAGGAAGATCG
g.s.-outer-pr	738	GCATCGTAACCGTGCATCTG
dG-anchor-p	740	CTACTACTACTAGGCCACGCGTCGACTAGTACGGGGGGGGGG
dG-sec-acho primer	741	CTACTACTACTAGGCCACGCGTCGACTAGTAC
beta-geo_R4	516	TCAGGCTGCGCAACTGTTGG

Differentiation of parthenogenetic ES cell clones. ES cells were dissociated using trypsin and EB formation was induced using the hanging drop technique (1000cells/20uL drop) in ES cell media without LIF. After 3 days, the EBs were collected and transferred to non-treated petri dishes for further differentiation. On day 5 and day 12, EBs were harvested and RNA was isolated using TRIzol Reagent (Invitrogen) according to the manufacturers protocol. 1µg RNA was reverse transcribed into cDNA using the iScript cDNA Synthesis Kit (Bio-Rad). Subsequently, qRT-PCR was performed with the GoTaq qPCR Master Mix (Promega, A6001)

using a Bio-Rad CFX384 Connect Real-Time PCR Detection System. The following primers (5' -> 3') were used for qRT-PCR:

Oct4: FWD CCTACAGCAGATCACTCACATCGCC

REV CCTGTAGCCTCATACTCTTCTCGTTGG

Klf4: FWD GTGCCCCGACTAACCGTTG

REV GTCGTTGAACTCCTCGGTCT

Nanog: FWD CAGGAGTTTGAGGGTAGCTC

REV CGGTTCATCATGGTACAGTC

Keratin 18: FWD TTGCCGCCGATGACTTTAGA

REV GGATGTCGCTCTCCACAGAC

Nkx2-5: FWD CCCAAGTGCTCTCCTGCTTT

REV TTTTTATCCGCCCGAGGGTC

Sox17: FWD GAATCCAACCAGCCCACTGA

REV TAGGGAAGACCCATCTCGGG

To assess protein expression of pluripotency genes during differentiation, the haploid ESCs were cultured in ES cell Media without LIF in the presence of 500nM all trans-Retinoic acid (Sigma, R2625) and the media was replaced every other day. At indicated time points the cells were fixed in 4% PFA and stained for Oct4 (BD Bioscience, 611202), Nanog (R&D Systems, AF2729) and Sox2 (R&D Systems, AF2018), counterstained with DAPI and imaged with a Zeiss LSM 780.

DNA content analysis and growth curves. Haploid AN3-12 cells were routinely passaged and fixed at indicated time points in ice cold 85% Ethanol and stored at -20. For analysis of DNA content, cells were stained with 10µg/mL Hoechst33342 (Sigma B2261) for 30min on ice, washed and analysed on a BD Fortessa. Viable cells were counted using a Countess II FL (Thermo Fisher) according to the manufacturer's protocol.

Keratinocyte differentiation. Embryoid bodies (EBs) from 2000 cells were generated using the hanging drop technique. After 3 days, 30-50 EBs were transferred to an Ultra-Low attachment 6 well plate and stimulated with 1µM retinoic acid for 3 more days. Next, EBs were transferred onto CollagenIV-coated dishes in ESCM without LIF medium supplemented with 25ng/ml hBMP-4 (R&D Systems). After 3 more days, the cells were incubated with CnT-07 keratinocyte medium (CELLnTEC) and cultured for 6 more days. Keratin14 positive cell population was enriched by using the rapid adherence to CollagenIV: trypsinized cells (0.25% Trypsin-EDTA) were plated onto CollagenIV-coated wells and non-adherent cells were washed off after 15min incubation at room temperature.

Reversion of gene trap cassettes. To reverse gene trap cassettes, ES cells were either stably infected or transiently transfected cells with a Cre- or Flpo-expressing plasmid also containing a fluorescent marker, e.g. EGFP or DsRed, using Lipofectamine 2000 (Invitrogen), Amaxa nucleofector (program A-13), or viral delivery as outlined above. If single cell derived clones were desired, cells were FACS-sorted for $Cherry^{+}$ or GFP^{+} cells 24-48h post lipofection/nucleofection and 1000 cells were then seeded on 15 cm tissue culture dishes. On day 10 post sorting, colonies were picked and used for further analysis. For stable integration

of Cre and the color marker upon infection, PlatinumE cells (Cell Biolabs) were transfected with the respective plasmids and ES cells subsequently infected. 30h post infection, cells were puromycin selected for further 72 hours. As controls, GFP-puro infected cells were selected and used. Inversion of the splice acceptor was determined by PCR on genomic DNA. The PCR was performed with three primers on ES cell lysates. Dependent on the orientation, either a fragment binding the 1st forward primer or a fragment binding the inverse forward primer were amplified. Upon first inversion, PCR fragments are larger in length. The second inversion results in a size reduction below the original orientation. The following specific PCR primers were used for Retro, Lenti, *Tol2*-EGT: EGT 1st fwd: CGACCTCGAGTACCACCACACT; EGT inv fwd: AAACGACGGGATCCGCCATGTCA*;* EGT com rev: TATCCAGCCCTCACTCCTTCTCT. Expected bands *for Retro, Lenti, Tol2*-EGT are: EGT 1st fwd/GT com rev: 343bps; EGT inv fwd/GT com rev: 443bps. The PCR primers for Tol2-pA were: Tol2 1st fwd: TGGGTTCAAGCGATTCTCCTGCCTCA; Tol2 inv fwd: AGATAGGCACCCAGGGTGATGCAAGCTC; *Tol2* com rev: CCGATCCATCCATCGCATATTTGGGA. Expected bands for *Tol2*-pA: polyA trap 1st fwd /GT com rev: 326bps; polyA trap inv fwd/GT com rev: 439bps.

RV-A1a infections. For virus production, Hela, cells at a confluency of 70-80%, were washed with PBS and infected with RV-A1a in infection medium (DMEM, 2% heat inactivated FCS, 1% Penicillin/Streptomycin, 1% L-glutamine, 30mM $MgCl_2$). After an incubation for 24 hrs, the virus was released from the cells by three freeze/thawing cycles and subsequently cellular debris was removed by centrifugation and filtration (4,000 rpm for 10 min at 4°C). For the genome-wide screen, retrovirally mutagenized haploid ES cells were seeded in 15cm

dishes ($1x10^7$ cells per dish), and 5 hrs post seeding cells were infected with RV-A1a at a MIO of 5 in a 1:1 mixture of Hela infection medium:ESCM ($MgCl_2$ and LiF were additionally added to balance the concentrations). For control mock infections, the Hela infection medium/ESCM mixture was used. On day 21 post infection, cells were pooled, and further processed for isolation of genomic DNA or frozen. Validation in ES cells was done using Haplobank clones (*Ldlr*: clone 1=238F02, clone 2=374A09, clone 3=1031A04; *Pla2G16*: clone 1=371H01, clone 2=588G11, clone 3 = 917E06). HEK293T cells from an institute resource were tested for neomycin resistence to confirm cell identity. They were infected with GFP, empty guide, and Cas9 containing control, or puro selection marker, guide and Cas9 containing control vectors. Subsequent to selection, CRISPR/Cas9 edited cells were mixed with GFP positive cells. These cell mixtures were monitored in the presence and absence of RV-A1a for shifts in ratios of GFP expressing cells using FACS (BD LSRFortessa and FlowJo). Arachidonic acid, Acetylsalicylic acid, Ibuprofen, Indomethacin, Diclofenac, and Celecoxib (all from Sigma Aldrich) in DMSO were added to the cells 2h prior to RV-A1a infection at the indicated concentrations. For interaction studies with Pla2g16 and Ldlr in mouse ES cells, the following drugs were used: Acetylsalicylic acid: 1mM, Archidonic acid: 10μM, Celecoxib: 910nM, Diclofenac: 380nM, DMSO (vehicle): 0.1%, Ibuprofen: 72μM, Indomethacin: 10μM.

The following guide sequences were used for mutagenesis:

GFP sg GAGCTGGACGGCGACGTAAA
LDLR sg1 TCAGACCGGGACTGCTTGGA
LDLR sg2 CTGTTGCACTGGAAGCTGGC
LDLR sg3 GCTGTTGCACTGGAAGCTGG
LDLR sg4 GGAGCTGTTGCACTGGAAGC
PLA2G16 sg1 GAAGGAATTGCTGTATGATG

PLA2G16 sg2 CCTGCAGCAAAATCATCCAG
PLA2G16 sg3 CTATGTTGGCGATGGATATG
PLA2G16 sg4 CGCTGGATGATTTTGCTGCA
Pla2g16 sgRNA1: TTGCTTCTGTTTCTTGTTTC
Pla2g16 sgRNA2: CTGAATGACTGCCCAGTTTT

Blood vessel sprouting assays. ES cells were trypsinized and seeded at 9600 cells/EB in a 96 well low attachment plate (Sumitomo Bakelite, Prime Surface-U). After incubation for 14 days in ESC Media without LIF in the presence of puromycin, the EBs were washed in ESC w/o LIF Media, embedded in 3D Collagen I gels and stimulated with 50 ng/mL VEGF-A (in house production) as previously described[32]. The first vascular outgrowth was observed after around 7 days and capillary like networks were analyzed 2 weeks after initial embedding of the EBs into the Collagen I matrix. For tip cell competition assays, mosaic EBs were generated from GFP and mCherry-Cre positive haploid ES cells. Briefly, Haploid ES cells were infected with viral supernatants harvested from Plat-E cells expressing GFP-Puro or mCherry-Cre-Puro. 48 hours after infection the cells were treated with 1 0µg/mL puromycin (Invivogen) to select for infected cells. Infected cells were FACS sorted for the 5% highest GFP/mCherry expressing cells, constantly kept under puromycin selection pressure to avoid reporter silencing, and then mixed at a 1:1 ratio to generate mosaic EBs**.** The following clones from the Haplobank were used to generate blood vessel organoids: Myst3 00235|A10; Myst3 00564|C06; Prkcz 00355|E11; Syt16 00160|B12; Tesc 00279|B08; Gfod1 00284|F11; Enpp3 00455|E06; Enpp3 00584|E08; Enpp3 00535|F07; Grin2b 00335|G12; Grin2b 00858|G05; Plcb1 00362|H01; Plcb1 00367|H09; Ndufs4 00318|D02; Ndufs4 10132|G01; Gja1 00345|A03; Tfpi 00281|H03; F2r 00351|E05; Pcsk6 00200|D09; Abcg2 00385|H12; Prkch 00102|C03; Igfr1 00535|B11; Entpd1 00588|B04; F3 00451|B01; Igf1 00455|D06;

Plaur 00582|C11; Hckl 00449|E02; Mapt 00569|D06; Dlg2 00339|A01; Tnfrsf1a 00112|H07; Smarca1 00182|D04; Cdh13 00255|G04; Ets2 00341|E06; Tgfbr2 00583|D04; Gabrb3 00578|B04; Abcb1a 00500|A06; Bag3 00338|G02.

Immunofluorescence. To visualize the vascular outgrowths, blood vessels were fixed with 4 % PFA for 20 min at room temperature and blocked with 3 % FBS, 1 % BSA, 0.5%Triton, and 0.5% Tween for 2 hrs. Biotinylated GSL Isolectin B4 (IB4; Vectorlabs 1201), anti-CD31 (BD Pharmingen AB9498), and anti-Collagen IV (Millipore AB765P) were applied over night at 4°C. After washing with PBS-T (0.05 % Tween), the organoids were incubated with streptavidin-Alexa488, anti-rat-Alexa488, or anti-rabbit-Alexa647 secondary antibodies. Before mounting with DAKO mounting media under a stereomicroscope, organoids were subjected to 1 μg/mL DAPI solution to image nuclei. A Zeiss LSM 780 was used to image the entire vascular outgrowth of every single organoid using tile scanning option and recording of z-stacks.

Confirmation of *Notch1* gene trap flipping by PCR and Western blot. DNA of stably GFP or mCherry-Cre infected cells was isolated using Phenol:Choloroform:isoamyl alcohol (25:24:1) (Sigma Aldrich) and Phase Lock Gel Heavy tubes (5prime). The following primer sequences were used for PCR: F1st: CTTCTGAGACGGAAAGAACCAGC, Finv: AAACGACGGGATCCGCCATGTCA Rev: TATCCAGCCCTCACTCCTTCTCT. The pair F1st+Rev amplifies a fragment of the original, unflipped gene trap and Finv+Rev amplifies the Cre recombined, flipped gene trap fragment. Western Blotting was performed using standard protocols. Blotted membranes were blocked in 3%BSA and subsequently incubated with

antibodies to detect cleaved Notch1 (CST D3B8) and β-actin (Sigma A5316) over-night. After washing in TBS-T (0.1%Tween) and incubating with HRP-linked secondary antibodies, blots were incubated in ECL solution (Pierce) for visualization.

Teratoma assays. ES cells were dissociated and resuspended in growth factor reduced Matrigel (Corning, 356231) and kept on ice. 10^6 cells were injected into both flanks of 8-12 weeks old MF1 *nu/nu* mice. Both, female and male mice were used. For generation of a chimaeric vasculature, $2x10^6$ ES cells were cultured in the presence of puromycin to avoid reporter silencing, in 15cm Petri dishes and mosaic EBs (GFP:mCherry 1:1) were generated. After 8 days of cultivation, the EBs were washed, resuspended in 200ml growth factor reduced Matrigel (BD 356231), and injected subcutaneously into the flanks of immunocompromised MF1 *nu/nu* mice using a 18G needle. Teratomas were analyzed 3-4 weeks after injection. To this end, teratomas were either processed for immunohistochemistry or cells were isolated for FACS analysis. H&E slides were reviewed with a Zeiss Axioskop 2 MOT microscope (Carl Zeiss Microscopy) and subsequently digitized with the Pannoramic Flash III (Adimec-Q-12-A-180Fc camera) automated slide scanner (3D Histech). For immunohistochemistry, teratomas were fixed with 4 %PFA over-night and subsequently incubated in 20 % sucrose for around 24 hrs before embedding into O.C.T and processing the tissue to 60-80mm thick sections using a cryostat. For visualization, sections were incubated with anti-GFP (Abcam, AB111258), anti-mCherry (Abcam, AB125096), and biotinylated-IB4 (Vectorlabs, #1201) over night. After washing and incubation with anti-goat-Alexa488, anti-mouse-Alexa555 and streptavidin-Alexa633, respectively, the specimens were imaged using a Zeiss LSM780 microscope. For FACS analysis, teratomas

were cut into 1mm pieces and subjected to enzymatic digestion (Collagenase Type IV 300U/mL, Worthingon; Dispase 0.25U/mL, Gibco; DNAse 7.5 μγ/mL, Qiagen). Following a 1 hr incubation at 37°C on a shaker, the cell suspension was pipetted through a 70 μm cell strainer, washed 1x with DMEM before centrifugation at 400 x g for 10 min. Single cells were incubated for 45min with an anti-mouse VE-Cadherin-APC antibody (ebioscience 17-1441-80) on ice. A ten min DAPI wash was performed to exclude dead cells from the FACS analysis. All animal studies were approved according to Austrian and EU legislature.

Retina analyses. Retinas were harvested and prepared from mice as previously described[35]. The Gja1/Connexin43 inhibitory peptide (GAP26 - VCYDKSFPISHVR) and the scrambled control peptide (sGAP26-YSIVCKPHVFDRS) were synthesized by PSL GmbH, Germany) and injected i.v. in neonatal C57BL/6 mice as previously described[36,48] at 2μg/mouse. Whole mount retinas were immunostained for IB4 (Vectorlabs, #1201) or Gja1/Connexin43 (abcam AB11370) expression, incubated with the appropriate secondary antibodies (see above), and imaged using a LSM 780 microscope.

Statistics. All values in the paper are means ± SEM or SD. Comparisons between groups were made by Student's t-test or 2-way ANOVA using GraphPad Prism (GraphPad Software, San Diego, CA) or R statistical software. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. $P < 0.05$ was accepted as statistically significant.

Data availability. All genomic data have been deposited to NCBI Expression Omnibus and are accessible through the GEO accession number GSE84090 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=qrgxmuaindmpjah&acc=GSE84090). All material, protocols and cell lines can be obtained via the website www.haplobank.at as well as via Nature Protocol Exchange (Haplobank methods collection).

Supplementary References

38. Elling, U. *et al.* Haplobank methods collection. *Protoc. Exch.* (2017). https://www.nature.com/protocolexchange/protocols/6135/edit/preview
39. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).
40. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
41. Jentsch, I., Adler, I. D., Carter, N. P. & Speicher, M. R. Karyotyping mouse chromosomes by multiplex-FISH (M-FISH). *Chromosom. Res.* **9,** 211–214 (2001).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
43. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
44. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* **12,** R22 (2011).
45. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28,** 511–515 (2010).
46. Mohn, F. *et al.* Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Mol. Cell* **30,** 755–766 (2008).
47. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **2015,** 21.29.1-21.29.9 (2015).
48. Gombash Lampe, S. E., Kaspar, B. K. & Foust, K. D. Intravenous injections in neonatal mice. *J. Vis. Exp.* e52037 (2014). doi:10.3791/52037
49. Golczak, M. *et al.* Structural basis for the acyltransferase activity of lecithin:retinol acyltransferase-like proteins. *J. Biol. Chem.* **287,** 23790–23807 (2012).

Extended Data Figure legends

Extended Data Figure 1. Stem cell properties of the haploid subclone AN3-12.

e-f, Various parthenogenic cell lines derived from independent embryos from an outcross of 129/Sv and C57/B6 and thus containing different genomic background on different chromosomes were subjected to embryoid body formation by placing 1000 cells/hanging drop. We observed downregulation of pluripotency marker genes **(a)** and upregulation of markers from all three germ layers **(b)** in all cell lines assayed on day 0 (d0), day 5 (d5) and day 12 (d12). The HMSc2 subclone AN3-12 was chosen for further study based on its growth properties in serum/LIF and absence of feeders. Data is shown as individual data points of n=2 technical replicates together the mean ± SD of one representative experiment. **c,** Growth curve of AN3-12 in the presence and absence of LIF. Data are shown as individual data points and mean values of 3 biological replicates ± SD. **d,** FACS analysis of chromosome content of AN3-12 cells (in LIF, same experiment as shown in panel c) shows the decline of haploid (1n) from 35.5% to 24.9% during the 7day culture period. **e,** AN3-12 cells, cultured as in panel c, maintain a robust haploid population when analyzed on day 17 in ES cell medium despite rapid proliferation. **f,** Differentiation of AN3-12 cells into keratinocytes resulted in a near complete loss of haploid cells among the Keratin 14 (K14) positive population; ES cells stained with anti-K14 are shown as a negative specificity control (grey curve in the K14 histogram). **g,** Immunostaining of AN3-12 cells cultured in ES cell medium as well as time course of removal of LIF with addition of 500nM retinoic acid (analysed on the indicated days) shows downregulation of pluripotency markers Oct4, Nanog, as well as Sox2. DAPI is

shown as a nuclear counterstain. Bar graphs 50μm **h,** Histological examination of Teratomas analysed 25 days after injection of 10^6 cells subcutaneously. All three germ layers were present in 6 analyzed teratomas, representative H&E images are shown. Magnifications are indicated in each panel.

Extended Data Figure 2. Analysis of genome integrity.

a, M-FISH karyotypic analysis was performed on parental mouse haploid cells (AN3-12) to evaluate genomic stability. Randomly selected metaphases were karyotyped and examined by M-FISH and DAPI banding. Approximately, 200 metaphases from AN3-12 were counted for the diploid versus haploid frequency and 10 well-spread metaphases were fully karyotyped by M-FISH and DAPI-banding pattern. Images of a normal female diploid and haploid karyotype [19, X] are shown. Images were captured on a Zeiss AxioImager D1 fluorescent microscope equipped with narrow band-pass filters for DAPI, DEAC, FITC, CY3, TEXAS RED, and CY5. **b,** CNV (copy number variation) analysis of haploid AN3-13 cells by genome sequencing using IlluminaHiSeq2500. Mapped reads were analyzed relative to male genomes of parental mouse strains C57BL/6J and 129/Sv respectively, quotient to closer parental strain is shown. As expected, the X chromosome is overrepresented while the Y chromosome is absent. Regions of detected variation are highlighted with red boxes and shown below. Chromosome numbers are indicated. **c,** In AN3-12 haploid ES cells three very small deletions (on chromosomes 2, 10, and 12) and 1 duplication (on chromosome 13) were detectable as highlighted. **d,** Chromosomal distribution of SNP densities for in-house 129/SV and AN3-12 ES cells relative to in-house C57BL/6 are shown. Numbers of SNPs were calculated for all non-overlapping 100 kb windows across the mm10 C57BL/6J mouse

reference genome. SNP density in AN3-12 shows regions of high and low number of SNPs relative to the C57BL/6J genome, as expected for a haploid cell line derived from an F1 female between 129/SV and C57BL/6.

Extended Data Figure 3. Molecular characterization of mutagenesis vectors.

a, Schematic illustration of the universal NGS sequencing strategy. Optimized primer binding sites compatible with Illumina sequencing and 2 restriction enzymes with 4 base pair recognition sites were placed adjacent to the terminal elements (LTR, TR). An internal barcode of 32 bases with alternating weak and strong bases was inserted in a parallel cloning step. **b,** For mapping of integration sites, genomic DNA was amplified by inverse PCR to introduce adaptor sequences and the experimental index for NGS. Paired-end sequencing maps the genomic integration in the first read using a custom primer, the experimental index as well as the internal barcode using standard Illumina primers binding to the integrated complementary sequence. Barcode (BC) PCR was performed on genomic DNA. **c,** Meta-analysis of mutagen integrations around transcriptional start sites (TSS) (excluding the precise TSS site). In particular Tol2 and Retrovirus show a preference to integrate in proximity to the TSS. Retroviruses also frequently integrate into the promotor regions, while lentiviral integrations are typically located within the entire gene body. IPKM= Insertions per kilobase per million. The vectors used are described in Fig. 1 legend. **d,** Distribution of integration sites. Binning the number of integrations in genic and 2kb upstream regions per 10kb windows illustrates pronounced cold spots of mutagenesis using retroviral

mutagenesis, where one can observe bins devoid of integrations. **e,** Genomic region surrounding the *Gapdh* locus exemplifying the distributions of integrations. While retroviral integrations strongly cluster, Tol2 displays a more uniform distribution of integration sites. Tracks are + strand (top) and – strand (bottom) integration sites. Bar lengths indicate NGS read numbers, subsequent to iPCR. **f,** Heat map illustrating overlap of epigenetic histone marks with integrations of the indicated mutagens, normalized to peak size. Only retrovirus and *Tol2* integrations strongly correlate with DNA accessibility determined by ATACseq and active marks such as H3K4me3 and H3K27ac. In silico mutagenesis is shown as a control.

Extended Data Figure 4. Insertional preferences and generation of mutant ES cell library.

a, Correlation of integration probabilities (IPKM=insertions per kilobase per million) to expression level (mean log2 of FPKM). Strongest correlation is seen for lentiviral constructs as well as Retro-GT without osteopontin enhancer elements. All mutagenesis vectors are described in Fig. 1 and methods. **b,** 5'RACE on a set of pooled clones with confirmed antisense integration sites revealed multiple spurious transcription initiation sites in the intronic part of the gene trap vector around the lox site, but we failed to detect spliced transcripts. Transcriptional initiation within the lox5171 site is highlighted. Red labelled sequence is marking polyGs used for 5' tailing. **c,** Intersection of integration sites of the indicated mutagenesis vectors (see Fig. 1) with genomic features. Coding sequences (CDS), 5' and 3' untranslated regions (5'UTR and 3'UTR), 1st Intron, all other introns excluding the first intron (Intron), non-coding exons (ncExon), upstream regions (defined as 2kb upstream of transcriptional start site), and intergenic regions are indicated. Mutagenesis by *Piggyback*

transposons as well as *in silico* random mutagenesis and ATACseq results are shown for comparison. **d,** Schematic work flow for generation of the mutant haploid ES cell library. Single cell derived clones were manually picked 10 or 11 days after seeding, expanded in 96 well plates, and either frozen in quadruplicates or further processed for mapping of the integration sites. **e,** Schematic illustration of the first step of 4D pooling. Each plate was pooled into the respective slice tray as well as a master-plate, uniting identical well coordinates of all plates. **f,** Schematic illustration of the second step of 4D pooling. Each master-plate was pooled into a master tower pool, a plate with lamella uniting columns, and a plate with lamella uniting columns, thereby generating pools for rows and columns over all samples. **g,** 4-Dimensional pooling of 9600 clones in 8 rows, 12 columns, 10 slices, and 10 towers resulting in 40 pools. Subsequent to iPCR to introduce experimental indices, pools were combined and deep sequenced. Amplification of internal barcodes confirmed clonal identity and mapping in 4 dimensions. All mapped clones were deposited to Haplobank (www.haplobank.at).

Extended Data Figure 5. Numbers of independent gene trap clones and intragenic distribution.

a, Numbers of independent available cell lines, carrying a single integration per cell, per gene. For about 37% (RefSeq) to 38% (Ensembl) of genes targeted, there is one gene trap clone available (5'UTR, Intron, or coding sequence), whereas about 18% of genes are targeted in two independent clones, and for ~ 43% of genes 3 or more independent clones are available. **b,** 24.8% (RefSeq) to 26.8% (Ensembl) of genes are represented by a single cell line if one takes all clones into account and about 40% of genes are hit in more 3 or more clones. **c,**

Separation of all gene traps combined into biotypes in single integration Haplobank clones. Antisense and intergenic insertions are observed in all systems, in particular for enhanced gene trap vectors. **d,** To map the integration sites of our Haplobank clones to the ORFs (open reading frames) of the respective genes dissected ORFs into 5% intervals and annotated integration sites in introns and exons relative to the position within the ORF. All mutagenesis systems (see Fig. 1) show a strong bias towards transcript truncation proximal to the 5'end of the ORFs and are thus predicted to result in loss of function alleles. We defined integrations in the anterior 50% of coding sequence (green bars) as optimal for a gene trap allele; these clones are highlighted by a yellow star on the Haplobank homepage.

Extended Data Figure 6. Interaction of Pla2g16 with Cox inhibitors in mouse ES cells.

a, Titration series of the indicated Cox inhibitors in the presence and absence of rhinovirus (RV-A1a) in mouse ES cells. No protective effect of inhibition of prostaglandin biosynthesis was detected at non-toxic concentrations. Since ES cells do not generate infectious RV-A1a efficiently, conditioned supernatant containing RV-A1a was added daily. Data is shown as individual data points and mean values. **b,** Haplobank ES cell clones harboring mutations in *Ldlr* and *Pla2g16*, respectively, were mixed as sister cells in sense (red) and antisense (green) orientation labelled by GFP and mCherry. Subsequently, cells were cultured in the presence and absence of rhinovirus RV-A1a for 4 days and ratios were then quantified using FACS. Selection pressure for loss of Ldlr and Pla2g16 was not affected by inhibition of Cox. Data is shown as mean of 3 biological replicates, error bars represent SD.

Extended Data Figure 7. Interactions of PLA2G16 with Cox inhibitors in human HEK293T cells and domain mapping.

a, RV-A1a exposure causes cell death in HEK293T cells in a dose dependent manner. Cell viability was quantified 3 days after infection by Alamar blue cell viability measurement. Data is shown as individual data points of 5 biological replicates and mean values. **b,** Titration series for ibuprofen and indomethacin treatment in the presence and absence of rhinovirus (RV-A1a) in human embryonic kidney HEK293T cells. Protective effects of ibuprofen and indomethacin were detected at high concentration. Cell viability was quantified 2.5 days after infection by Alamar blue cell viability measurement. Data is shown as individual data points of 4 biological replicates and mean values. **c,** Competitive growth assays in HEK293T cells. Cells containing sgRNAs targeting PLA2G16 did not show a growth difference in the absence of RV-A1a or when treated with indomethacin at 100μM, but were significantly enriched when challenged with RV-A1a, indicting preferential survival. By contrast, control guide treated cells did not show growth advantages at any experimental condition. Data are shown as individual data points and mean values +/- SD of biological triplicates analysed on days 0, 3, 6, and 10 after RV-A1a exposure. **d**, Scheme of Pla2g16 domains. The enzymatic center of Pla2g16 is located in the cytoplasm (green); an alpha-helix in the transmembrane domain (yellow) connects it to a short vesicular domain (blue), located in endosomes[49]. **e,** Design of CRISPR sgRNAs targeting the mRNA regions encoding the vesicular domain of 7 amino acids (sgRNA1) and the 3'UTR (sgRNA2) in haploid ES cells to test essentiality of these domains in RV-A1a infections. **f,** Cells carrying sgRNA2 showed editing in only 1/12 cases, but upon

selection with RV-A1a were enriched for deletions within the vesicular domain. For sgRNA2, all mapped deletions in control cells only affected the 3'UTR, where the expected Cas9 cuts occur; upon RV-A1a exposure, the majority of observed deletions affected the transmembrane domain, the vesicular domain, and in some cases even extended into the cytoplasmatic region. Color codes: Grey: deletion; Red: alternative reading frame and insertions.

Extended Data Figure 8. Blood vessel sprouting in *Notch1* mutant ES cells and candidate tip cell genes.

a, Assessment of 4 independent *Notch1* targeted clones from Haplobank. The locations of the integrations are shown: 2 anti-sense (as) clones marked by green triangles, 1 sense (s) clone marked by the red triangle, and one clone with an upstream (as-up) integration (blue triangle). Flipping of the gene traps upon Cre infection is shown by PCR in the middle panel. Loss of Notch1 protein (intracellular domain, ICD) expression (clones A4, H7), and re-expression (clone D5) upon Cre recombination are shown by Western blot (lower panel). β-actin is shown as a loading control. WT, parental clone without any gene trap integration. Uncropped blots are shown in Supplementary Figure 1. **b,** Notch1 inactivation leads to a hyper-sprouting phenotype. Note the advanced progression and increased density of the vascular networks upon *Notch1* deletion (sense clone) compared to anti-sense sister cells (upper panels – bright field images, lower panels IB4 immunostaining to mark endothelial cells). Scale bars 500μm. **c,** Angiogenic sprouting is not affected when the gene trap is located 1500bp upstream of the *Notch1* gene (A2 clone). For molecular characterization of the A2

clone see Extended Data Fig. 12. GFP$^+$ and Cre-reverted mCherry$^+$ sister cells were analyzed in 3D blood vessel organoid cultures. Bright field images are shown. Scale bars 500 µm. **d,** Differentially expressed genes in endothelial tip cells versus stalk cells from two published datasets[30,31] in the murine retina were filtered for genes that have also been associated (Ingenuity pathway analysis) with candidate genes/pathways for vascular diseases in humans. Scatterplot showing the frequency of independent associations of "tip cell genes" with various human vascular diseases. Genes available at Haplobank at the beginning of the project were chosen for functional analysis in the 3D organoids. For most of the listed candidate genes, there were no functional vascular data available. **e,** Quantification of IB4 positive vascular structures from the indicated sister clones carrying sense and repaired antisense integrations. Clones were classified according to their sprouting capacity from low (hypo-sprouting) to high (hyper-sprouting).Data are shown as individual data points from a minimum of n=3 independent experiments for each sense/antisense sister clone combination together with the mean values ± SEM. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (Two tailed students t-test). **f,** Different clones with independent integrations in the same gene showed reproducible phenotypes in sprouting angiogenesis. Vascular outgrowths were stained for endothelial specific IB4 expression, number of vessels counted and normalized to the respective anti-sense sister clones. Data are shown as individual data points from a minimum of n=3 independent experiments for each sense/antisense sister clone combination together with the mean values ± SEM. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (Two tailed students t-test).

Extended Data Figure 9. Sprouting angiogenesis in revertible sister clones.

Representative images of the indicated sense (S) and anti-sense (AS) sister clones. IB4 was used to mark endothelial cells. GFP or mCherry expression indicates the respective flipped gene traps. Note that some sense clones are GFP$^+$ while others are mCherry$^+$; this is due to the original orientation of the integration in sense or anti-sense, which was then reverted by the mCherry-Cre expressing virus. Scale bars for all images is 500μm. For quantification of data see Fig. 3e.

Extended Data Figure 10. Generation of a chimeric vasculature *in vivo* and Gja1/Connexin43 localization to tip cells

a, Representative fluorescence image of a haploid ES cell-derived teratoma stained for endothelial specific IB4. Endothelial cells arising from haploid ES cells are positive for mCherry and IB4 (yellow), whereas host endothelial cells are only positive for IB4 and appear green. Scale bars 50 μm. **b,** Representative FACS analysis of teratomas following injection of chimaeric EBs into immunocompromised mice. *Myst3* anti-sense (mCherry$^+$) and sense (GFP$^+$) sister clones were mixed at a 1:1 ratio. VE-Cadherin-negative non-endothelial cells were also determined within the teratomas. **c**, Parental haploid ES cells stably expressing GFP or mCherry-Cre were assessed for their ability to generate IB4$^+$ vascular structures in the presence of VEGF-A. The number and ratios of IB4$^+$ vessels per organoid were not apparently different between GFP and mCherry-Cre expressing cells. Scale bars 500μm. Data are shown as individual data points from n=4 independent experiments mean values ± SEM. P=0.207 (Two tailed students t-test). **d,** GFP and mCherry-Cre expressing parental haploid ES cells contribute equally to tip cells (49.2 % GFP$^+$; 50.8 % mCherry-Cre$^+$) in 1:1 mixed mosaic cultures. Data are shown as individual

data points from of n=4 independent experiments together with the mean values ± SEM. P=0.823 (Two tailed students t-test). **e,** Localization of Gja1/Connexin43 protein in the murine retina at postnatal day 6 (P6). Endothelial cells are marked by IB4 staining. At the angiogenic front, Gja1/Connexin43 expression is found in endothelial cells, primarily localized at tip cells (arrows). Scale bars 50µm. **f,** Retinas were stained for Gja1/Connexin43 protein expression and the endothelial marker IB4 to visualize the vascular networks on postnatal day 6 (P6). Note punctate pattern of Gja1/Connexin43 adjacent to the $IB4^+$ vessels, suggestive of Gja1/Connexin43 expression in perivascular cells. Scale bars 50 µm. **g,** Gja1/Connexin43 protein predominantly localizes to the tip cells (arrows) in the 3D blood vessels. Vessels are marked by CD31 immunostaining and counterstained by DAPI. Bar graph indicates percentages of vessels with highest Gja1/Connexin43 expression in the tip cell; Data is shown as individual data points of 8 independent EBs together with the mean values ± SD of vessels. Scale bars upper panels 20µm, lower panels 10µm.