Mammalian Genome (2021) 32:1–11
https://doi.org/10.1007/s00335-020-09853-1

Ratpost: a searchable database of protein-inactivating sequence variations in 40 sequenced rat-inbred strains

Steven Timmermans[1,2] · **Claude Libert**[1,2]

Received: 15 September 2020 / Accepted: 18 November 2020 / Published online: 22 January 2021

Abstract

Rat-inbred strains are essential as scientific tools. We have analyzed the publicly available genome sequences of 40 rat-inbred strains and provide an overview of sequence variations leading to amino acid changes in protein-coding genes, premature STOP codons or loss of STOP codons, and short in-frame insertions and deletions of all protein-coding genes across all these inbred lines. We provide an overview of the predicted impact on protein function of all these affected proteins in the database, by comparing their sequence with the sequences of the rat reference strain BN/SsNHsdMcwi. We also investigate the flaws of the protein-coding sequences of this reference strain itself, by comparing them with a consensus genome. These data can be retrieved via a searchable website (Ratpost.be) and allow a global, better interpretation of genetic background effects and a source of naturally defective alleles in these 40 sequenced classical and high-priority rat-inbred strains.

Introduction

The use of the rat (NCBI Taxon ID 10116, *Rattus norvegicus*), as a model organism in research has a long history and started in the nineteenth century. The rat is considered to be the first organism to have been domesticated for scientific research purposes (Modlinska and Pisula 2020; Richter 1959). While some sporadic experiments using rats were done in the first part of the nineteenth century, such as an experiment on fasting in 1829 (Hedrich 2000; Modlinska and Pisula 2020), the systematic use of rats has started in 1856 (Philipeaux 1856), with studies on the impact of adrenalectomy (Philipeaux 1856) followed by studies on nutrition (Savory 1863), behavior, neurology, and others (Bergman et al. 2000; Watson 1914; Jonckers et al. 2011; Logan 1999). In some fields, rats outperform mice, mainly because of their bigger size, allowing larger biopsies and more repetitive (blood) samples to be taken, higher resolution imaging (e.g., in cancer therapy Bergman et al. 2000) and better precision in surgery (Jonckers et al. 2011; Modlinska and Pisula 2020). In cardiovascular research and neurobiology, rats are often preferred above mice, based on size and the amount of already available datasets and papers (Ellenbroek and Youn 2016; Leong et al. 2015). On the other hand, rat research is more expensive than mouse work, as rats require larger cages and more space. Like in the mouse field, rat research is mostly performed in rat-inbred strains, but outbred lines, typically yielding individuals with less defined genetic background, are also used (Sharp and Villano 2013).

The laboratory rat is the second most used animal model organism of all species; for example, in the European union rats accounted for 12.2% of all animals used in 2017, preceded only by the mouse (60.8%). This popularity was one of the reasons why the rat was selected to have its genome sequenced with priority, which was possible via a whole genome shotgun sequencing approach. The first draft of the rat reference genome was published in 2004 by the rat genome sequencing consortium (Gibbs et al. 2004). This genome was derived from the inbred Brown Norway strain (BN/SsNHsd), which was selected by the community and inbred another 13 generations at Medical College of Wisconsin (to BN/SsNHsdMcwi) before sequencing 2 females of this strain (Gibbs et al. 2004).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00335-020-09853-1.

✉ Claude Libert
claude.libert@irc.vib-ugent.be

[1] VIB-UGent Center for Inflammation Research, Ghent, Belgium

[2] Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

However, several rat strains/stocks were specifically bred to yield and investigate specific phenotypes or diseases. For inbred lines, well-known examples are the spontaneously hypertensive rat (SHR) strains that develop spontaneous hypertension (Conrad et al. 1995) or the Zucker Rat strain develops obesity (Obrosova et al. 2006). In outbred stocks, the outbred heterogeneous stock rat is one of the best known, created from 8 founder lines and used to study complex traits (Woods and Mott 2017) (e.g., depression disorder (Holl et al. 2018)). It is thus important that genetic variation between the strains is investigated in order to help explain the genetic components underlying these and other phenotypes. The sequencing of 28 rat (sub)strains by Atanur et al. (2013) and the sequencing of 40 strains (including resequencing and/or reanalysis of the previous 28) by Hermsen et al. (2015) encompass all the data that are currently available. These can be accessed at the rat genome database (RGD, Smith et al. 2020), where also additional information can be found concerning gene ontologies, QTLs, phenotypes, variants, pathways, as well as various tools to use this information (e.g., ontology enrichment, pathway browser).

While the rat sequence variations data are available in the RGD database, the data are provided to the user in a variant browser or as compressed vcf files. The information provided for coding genes is limited in the sense that only the nucleotide variation and the effect on the coding sequence (missense, nonsense) is given while there is no detailed overview over the consequences of the amino acid variation on the protein function, or a way to see the combined effect on the final amino acid sequence at a given position in a certain strain, compared to the reference strain amino acid. The latter can be important, as multiple mutations may affect a single codon, it was even shown that many of the nonsense variants actually have a second SNP that leaves a missense instead of a nonsense variant (Hermsen et al. 2015). However, the RGD does provide information about non-coding variants, where our tool provides only details about protein-coding variants. This makes the RGD a highly complementary resource to the tool described in this work.

Since a single amino acid change can have impact on protein function, ranging from no impact to a mild change or ruining the proteins function entirely, the coupling of amino acid variations to predicted function changes is essential. We have addressed this problem in the past, studying all amino acid changes in all 42 sequenced mouse inbred strains, and combined this information in a user-friendly database, called 'mousepost' (Timmermans et al. 2017). In the current study, we have focused our attention on the analysis of protein-coding variants in the 40 sequenced rat stains, as compared to the reference strain BN/SsNHsdMcwi. We have focused on high-confidence homozygous variants in inbred strains and report those contributing to changes in amino acid sequence, viewed on a per codon basis. Note that several of these are derived from outbred stocks such as the WKY/N strain from Wistar rats https://rgd.mcw.edu/. Our data allow thus to link phenotypes of rat strains with amino acid polymorphisms, as well as studying all variant versions of a given protein across all sequenced rat strains, or a genome-wide strain-to-strain comparison. Finally, we have also focused on the potential defects in protein-coding genes in the reference strain BN/SsNHsdMcwi, by comparing its genome to a consensus genome, defined by the other 40 sequenced strains.

Results and discussion

Classification of variants and comparison with the reference rat genome

The first draft assembly of the rat (*Rattus norvegicus*) genome was released in 2004 (Gibbs et al. 2004) and has been improved since. The reference genome is derived from the BN/SsNHsdMcwi inbred rat strain. The genome of 40 other high-priority rat-inbred strains has also been sequenced (Hermsen et al. 2015). The rat genome browser (RGD, https://rgd.mcw.edu/) is a rich database which contains the reference rat genome sequence, rat phenotype information, protein–protein interaction data, nucleotide variants, from more than 100 frequently used rat-inbred strains, and outbred stocks. Most of the research of today is preferentially performed in inbred lines, because of the stability in time and space of an inbred genome, and the identical background of individual animals within a certain inbred strain. Outbred animals are often used for more general research and behavioral studies. The identification of protein-coding variations among the rat strains is important for the correct interpretation of research or may serve as a useful tool for genetic research.

Since even a single nucleotide sequence polymorphism (SNP) variation in a protein-coding gene can have a severe effect on the protein function, via an amino acid change or the formation of a nonsense mutation STOP signal, we decided to transform the tabular lists of SNPs, and their positions in the rat genome, in protein-coding genes, into amino acids changes in the corresponding proteins, for each protein across all 40 sequenced rat strains, with the reference genome as a standard.

Sequence variations of a given rat-inbred strain compared to the reference strain are found in 'variant files' in the RGD. As a first step, we processed the variants files to retain only high-quality events. For each strain, coding genes and SNP/indel events were filtered as to only retain relevant genes (containing at least one nucleotide variant in that strain) and events overlapping with exons. Transcripts were classified exclusively in one of three classes: (1) stop gained (SG), which are sequence variations leading to a new STOP codon

(missense mutation), (2) stop lost (SL), which are nucleotide changes leading to a loss of a STOP codon and further translation of the mRNA and (3) non-synonymous variants (MUT), which are small insertions and deletions of maximally 200 bp and SNPs, as shown in Fig. 1. For each class, transcripts and all accompanying data and results are stored in a relational database for use in the web tool. An SG or SL variant takes precedence over the MUT class, so a gene containing an SG and any number of MUT events will have the label SG. These data also include splice variants, but only in the case of a mutation in the canonical splice donor or splice acceptor sites. Such sites are not specially annotated and are then seen as either intron retention or exon loss, and the variant CDS is constructed with these variations included. The final effect is most likely a frameshift and nonsense mutation, and these are included in the corresponding class (usually SG).

The impact of SG or SL variants was estimated by the length ratio, i.e., we assumed that the greater the deviation from the reference protein chain length, the more likely that there will be an impact on function. In general, a nonsense variant removes 46.6% of the total sequence (median value,

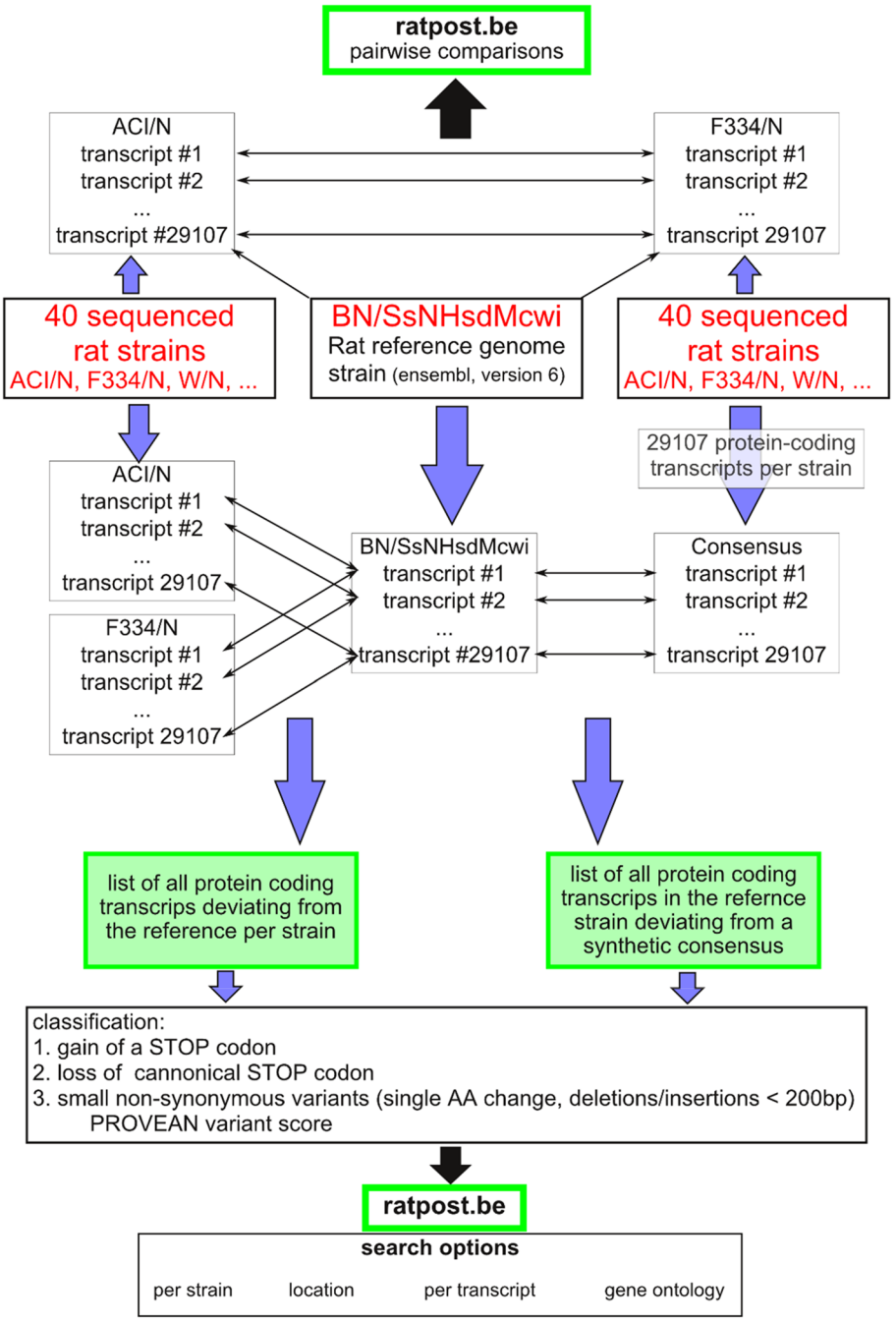

Fig. 1 Data use flowchart: data are obtained from ensemble (genome annotations) and the rat genome database (reference genome, sequence variants). Protein-coding transcripts are filtered from the genome annotation and their exon sequences extracted from the genome. Only transcripts with at least one exon containing a variant are kept. Variants are likewise filtered so only those overlapping an exon are kept. We perform three analyses: First, we compare the strain-specific sequences of each individual strain to the reference sequence. Second, we construct a synthetic consensus reference from all strain-specific sequences using a majority vote method (per position), and the reference sequence is compared to this consensus in order to obtain genes that can be considered to be deviant in the reference strain. For these two analyses, protein sequences are classified as stop gain, stop loss, or mutated (non-synonymous variant(s)). Potential effects of mutated variants are estimated with PROVEAN. Finally, we perform a pairwise comparison between each of the non-reference strains to obtain a list of all differences. All these data are made available in the website "ratpost.be" and can be accessed through strain and class-specific listing or searched by gene, location, or gene ontology term

mean: 47.2%), resulting in a high impact. For MUT variants, we applied the Protein Variant Effect Analyzer (PROVEAN) software to provide a score to a given variation (Choi et al. 2012). We are aware that the SL- and SG-classified transcripts may contain additional non-synonymous variants, but these were not analyzed further with PROVEAN. We made this decision by taking into consideration the additional information gained by performing a PROVEAN analysis compared to the time and the computational resources needed to obtain these scores. The variant sequences themselves are available in the web tool. The PROVEAN score indicates the chance that the variant amino acid will affect the protein function. The lower the score, the higher the chance of function loss. Based on previous papers from the authors of the PROVEAN software (Choi and Chan 2015a; Choi et al. 2012), we here considered PROVEAN scores lower than − 2.5 as leading to significant function loss. This is the point where the software shows the optimal balanced accuracy, i.e., where the sensitivity and specificity are both within acceptable limits. For PROVEAN, this means a sensitivity and specificity of about 80%. A higher specificity (90%) can be obtained by changing the score cut-off (− 4.1) at the cost of the sensitivity, which drops to 57%. We tested all 32,883 genes included in the ensemble genome annotation for the rat reference genome (41,078 transcripts). Our analysis showed that 12,172 distinct transcripts from 9715 genes have at least one natural variant in the included rat strains, which corresponds to 29.6% of the transcriptome or 29.5% of all genes. The total amount of different variants per strain and per class is shown in Table 1. As expected, the amount of non-synonymous variants (MUT) is the largest group. However, only a fraction of those (28.7%) are predicted to be potentially deleterious for protein function (PROVEAN score < − 2.5). The amount of nonsense (SG) variants in each strain compared to the reference strain is lower compared to the missense variants, with a maximum of 627 transcripts affected in any one strain (median: 558, IQR: 160). The amount of transcripts affected by a variant removing the stop codon is lower, with 67 or less transcripts affected per strain (median: 45, IQR: 1). While the number of non-synonymous variants is the highest, only 20–30% of those have a PROVEAN score of − 2.5 or lower, so the majority of them predicted not to influence protein function. The other BN strains ("BN/SsN", "BN-Lx/Cub", "BN-Lx/CubPrin"), which are closest related to the reference strain, display the lowest number of variants, as expected. In absolute numbers, the non-BN strains are all in the same range for amount of variant-containing transcripts (median(IQR): 599(43), 55(7), and 4971(518) for SG, SL, and MUT, respectively). In the 3 BN strains, the proportion of SG variants on the total is much higher than in the other non-BN strains. With an average of 278 transcripts that contain a stop-gain variants in the "BN/SsN", "BN-Lx/Cub", "BN-Lx/CubPrin" strains, 45.5% of all variant-containing transcripts in these strains fall in the SG category. For the non-BN strains, there are on average 532 stop-gain containing transcripts per strain, amounting to 9.5% of all variant-containing transcripts in each strain, on average. Many of the sequenced strains are also related closely to one another or are derived from a common stock; for example, the SHR strains. Compared to the reference strain, these have similar amounts of variants and many of these are shared between several or all these strains. In case of the SHR strains, 64% of all events (2035 of 3329) are shared between at least two and 23% between all of them (764 of 3329), as is illustrated in Fig. 2. Note that one of the SHR strains, SHR/OlaIpcvPrin has a larger amount of strain-specific variants (914) than the other three (43, 98, 139).

Data availability

All data displayed in Table 1 are available through the searchable *Ratpost* (*rat po*lymorphic *s*equence t*a*gs) web tool, which is available at ratpost.be. The database provides an overview table showing the number of SG, SL, and MUT variants per strain, with customizable cut-offs for length ratios and PROVEAN scores, and the numbers in the table are clickable. A listing of the strain-specific variants (versus the reference strain), divided by type, can also be obtained and exported in combination with various advanced filter options.

Ratpost provides several search functions that may assist with various questions. For example, (1) the tool provides the option to search for all variants of a particular protein-coding gene or transcript, across all 40 strains, or in a subset of these strains; (2) a search for variants, based on chromosomal location, can be helpful in the process of mapping and positional cloning of a trait; and (3) searching deviant protein-associated functions (by introducing a gene ontology term) is also possible. As a way of providing examples, which illustrate the power of the tool, we investigated several variants that are potentially interesting and/or relate to known (strain-specific) rat phenotypes. We provide three examples.

One of the best known variants in several mammalian species is the loss of function of tyrosinase (*Tyr*), resulting in an albino phenotype (Blaszczyk et al. 2005). In *Ratpost*, by introducing 'Tyr' as a search function, we recovered the previously identified *Tyr* variant (Blaszczyk et al. 2005) with a loss-of-function effect, the R299H mutation, which has a PROVEAN score of − 3.2 and is found in 32 strains. This is due to the fact that many strains that were sequenced are albino strains having the defect Tyr gene.

The Wistar outbred stock is well known as a model for diabetes since these rats develop diabetes spontaneously (Obrosova et al. 2006). Four inbred strains derived from

Table 1 Number of genes and transcripts per class

Strain	Stop gain (SG)		Stop loss (SL)		MUT		MUT (< − 2.5)	
	Genes	Transcripts	Genes	Transcripts	Genes	Transcripts	Genes	Transcripts
ACI/EurMcwi	371	421	45	51	3962	4916	1099	1296
ACI/N	528	604	58	64	4193	5210	1173	1397
BBDP/Wor	515	583	55	60	3974	4929	1191	1409
BN/SsN	193	224	24	27	180	222	43	47
BN-Lx/Cub	216	251	25	29	264	331	59	67
BN-Lx/CubPrin	315	360	37	41	272	348	58	67
BUF/N	318	360	43	47	3695	4582	1024	1216
DA/BklArbNsi	491	557	57	62	4048	5034	1164	1375
F334/N	343	383	43	46	3734	4638	1063	1251
F344/NCrl	480	545	54	58	4050	5026	1182	1390
F344/NHsd	493	558	54	58	4048	5024	1180	1391
FHH/EurMcwi	488	550	62	67	4217	5206	1241	1446
FHL/EurMcwi	502	564	56	63	4339	5380	1226	1452
GK/Ox	511	573	58	62	4083	5094	1236	1463
LE/Stm-Illumina	516	582	51	56	4102	5081	1224	1437
LE/Stm-SOLiD	327	374	36	40	3699	4622	1077	1275
LEW/Crl	478	539	48	52	3530	4374	1037	1220
LEW/NCrlBR	478	539	47	51	3542	4390	1025	1201
LH/MavRrrc	506	571	49	54	3870	4817	1147	1353
LL/MavRrrc	501	568	51	55	3935	4883	1164	1373
LN/MavRrrc	324	371	39	43	3617	4487	993	1191
M520/N	504	581	53	58	3778	4645	1127	1312
MHS/Gib	501	578	50	55	3877	4815	1161	1385
MNS/Gib	330	376	35	39	3695	4600	1053	1245
MR/N	517	586	52	57	4065	5037	1164	1370
SBH/Ygl	492	559	50	54	3984	4943	1142	1372
SBN/Ygl	550	627	54	60	4487	5570	1333	1570
SHR/NCrlPrin	535	606	51	57	4490	5575	1332	1562
SHR/NHsd	534	614	47	51	4422	5496	1303	1529
SHR/OlaIpcv	531	614	52	58	4464	5538	1333	1565
SHR/OlaIpcvPrin	478	543	46	49	4008	4971	1187	1395
SHRSP/Gla	510	585	48	52	4503	5576	1335	1568
SR/Jr	485	550	49	54	3940	4888	1167	1380
SS/Jr	481	553	48	52	3985	4940	1165	1371
SS/JrHsdMcwi	482	551	53	57	3984	4940	1152	1352
SUO/F344	469	528	61	65	4146	5111	1255	1446
WAG/Rij	492	564	50	56	3661	4502	1080	1257
WKY/Gla	317	372	36	37	4036	5045	1173	1383
WKY/N	540	613	53	58	4516	5607	1344	1576
WKY/NCrl	490	564	47	51	4298	5370	1245	1478
WKY/NHsd	500	577	49	53	4358	5437	1278	1509
WN/N	341	385	35	40	3786	4692	1062	1251

For 'MUT' the total number and the number with PROVEAN score < − 2.5 are shown

Wistar rats have been sequenced (WKY/N, WKY/Nctrl, WKY/NHsd, WKY/Gla). Searching for genetic variants in genes related to diabetes (by entering the GO term 'glucose metabolic process') provides *Gckr* (glucokinase regulator) with the variant R149C (PROVEAN score − 4.9) found in three out of four strains (WKY/N, WKY/Nctrl, WKY/NHsd) as a likely polymorphism related to the diabetes phenotype. It has been shown that both a genetic and an environmental component are involved in developing the disease (Liu et al. 2015; Stumvoll et al. 2005), and this gene may contribute to

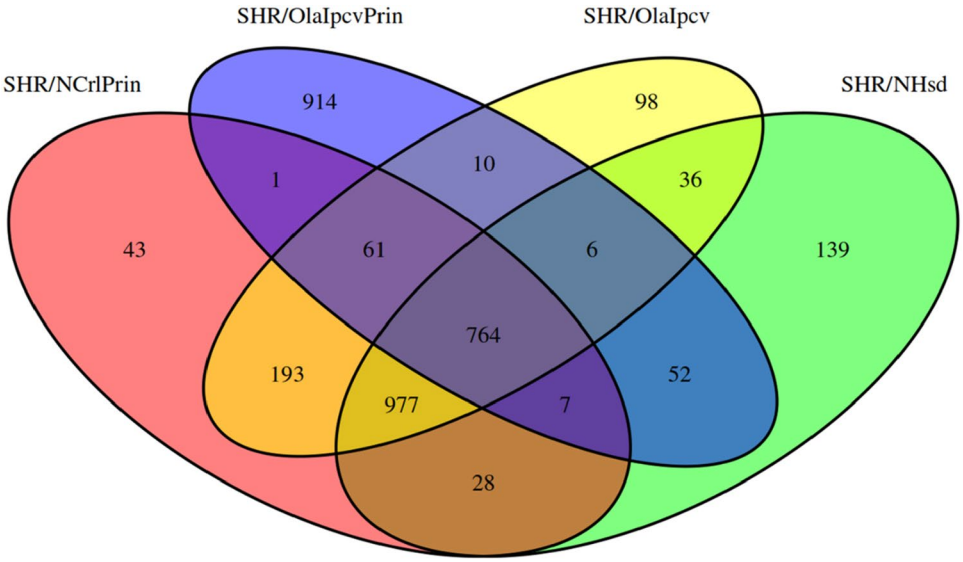

Fig. 2 Overlap between MUT events in all transcripts (PROVEAN < − 2.5) in the SHR strains

the former. The gene itself is a strong candidate for maturity onset diabetes of the young in humans (O'Leary et al. 2016). Another hit in this analysis is *Phip*, which has the GO function 'insulin receptor binding' (GO) and is involved in promotion of development and survival of pancreatic β cells (Podcheko et al. 2007). The encoded protein has a W388L variant, (with a PROVIAN score of − 10.7) found in the GK/Ox strain, a standard rat model for rat type 2 diabetes (Liu et al. 2015; Stumvoll et al. 2005).

The SHR strains develop spontaneous and severe hypertension (Conrad et al. 1995). This phenotype was linked to several genomic loci. When interrogating *Ratpost*, we found a few dozen genes at these locations with predicted deleterious variants present in these strains, and so we were not able to clearly identify a major candidate gene (Smith et al. 2020), but interestingly, we found that these strains also have a mutation in the gene *Mybphl* (W159L, PROVEAN of − 12.7). The protein (Myosine-binding protein H-like) plays a role in cardiac function (Barefield et al. 2017) and may sensitize these rats for developing the cardiac issues (heart failure) that they are known to develop as a result of their the hypertension (Trippodo and Frohlich 1981).

As a final example of the value of the database, in Supplemental Table, we provide the results of a search with the GO term "inflammation," which yield a variety of interesting deviant alleles present in several rat-inbred strains, e.g., in the genes *Tlr3*, *Tlr10*, *Tlr12,* or the Mineralocorticoid gene *Nr3c2*.

Variations in the reference and pairwise comparisons

As shown in Supplemental Figure, most variants are specific to only one strain, or to a rather small group of strains. However, it is clear that there are also some variants that are found in many strains, resulting in a situation where the *alternative* sequence is the one that occurs in the majority of strains that have been sequenced. This means that there are also genes where the reference strain is the one containing the *variant* or *defective* gene: similarly as the mouse reference strain C57BL/6J has been shown to harbor multiple variant (even defective) protein-coding genes compared to all other strains (Timmermans and Libert 2018), we categorized genes where the reference strain, BN/SsNHsdMcwi, contains a deviant allele. In *Ratpost*, we provide a way to query these variants, which are based on a comparison of the protein-coding sequences of the reference strain with a 'consensus rat genome': by using a simple majority-voting mechanism, the amino acid sequence that is supported by the most strains at each position of each protein is fixed and is considered a synthetic reference to which the BN/SsNHsdMcwi sequence is compared. Due to the nature of the data, the following considerations need to be taken into account: (1) as only a limited number of the available rat strains have been sequenced (40 plus the reference strain out of > 100 strains), the true consensus sequences cannot be known with certainty; (2) if the sample from the set of strains is biased, then the consensus sequence will be biased as well. This actually the case to some extent for the dataset: there are several closely related strains sequenced (SHR, WKy, and F344 substrains), it is not unlikely that if 1 SHR strain supports a certain variant, all strains will support this variant. However, as shown previously (Fig. 2), these related substrains still contain variants that are not shared by all of them. Overall, all the most common strains are represented in the dataset, while some bias cannot be excluded, we believe that this gives a high-quality overview of variation in commonly used inbred rat strains. For example, the defective mutant

Tyr gene is present in a majority of the sequenced rat strains (32 out of 41), and would be considered as the consensus sequence, so the wild-type *Tyr* sequence of BN/SsNHsd-Mcwi would wrongly be identified as deviant. These facts mean that, while potentially very informative, care needs to be taken when interpreting the results of this BN/SsNHsd-Mcwi versus the consensus reference. Such variants, with the exception of stop losses (these need to have the 3′ UTR added to detect the presence of a new stop codon in some cases), can be derived from the *Ratpost* database that was constructed as described in the previous section.

A list of reference-strain-specific variants can thus be obtained in *Ratpost*. Only variants where more than 50% of the strains differ from the reference are included in this analysis. We also provide an agreement score ($S = M1 \times M2$), which is a number between 0 and 1 that shows how many strains (1) agree with the consensus sequence (metric $M1 = C/N$, C: number of strain supporting consensus, N: total number of strains) and (2) how many other variants there are for that position that are not the consensus sequence and not the same as BN/SsNHsdMcwi (metric $M2 = 1 - B/N$; B: number of strains supporting the BN/SsNHsdMcwi sequence, N: total number of strains), the closer this number is to "1", the more unique the BN/SsNHsdMcwi sequence is. When performing this analysis, we find 411 transcripts where the BN/SsNHsdMcwi has a stop-loss variant compared to the consensus sequence and 49 where it has a stop-gain mutation compared to the consensus. For non-synonymous variants, we find a total of 5981 variants that differ in the reference rat strain compared to the synthetic consensus sequence. Assuming that the variants with a PROVEAN score of < -2.5 in the reference are negatively affecting the protein function, there are 1191 variants present that have a negative effect on protein function (out of 5981 in total). It is important to take into account the fact that the variants found in BN/SsNHsdMcwi compared to the consensus sequence across all strains are not always the correct ones, it is not because the large majority of the sequenced strains have a certain agreement sequence that this sequence is the functional one. These results also show that it is important to keep the genetic background of the used animals when using the rat reference genome, which contains its own set of defective alleles. Of the variants with low PROVEAN scores, there are 52 (in 37 transcripts) with an agreement score of $S = 1$, i.e., unique to the reference strain (see Table 2). One example is a variant unique to the reference that is the gene coding for proline dehydrogenase 1 (*Prodh1*) which has a G373C substitution with a PROVEAN score of − 8.3 as well as a second substitution: G312C with a score of − 8.6. While the reference strain has not been extensively tested in this regard, it is known that mutations in his gene in humans lead to Proline Dehydrogenase deficiency (Afenjar et al. 2007). Depending on the residual activity phenotypes can be very mild to severe, manifesting as hyperprolinemia with cognitive defects in severe cases (Afenjar et al. 2007; Perry et al. 1968). A second interesting example concerns the *Dclk1* gene, also specific to the reference strain, which has two amino acid substitutions compared to all other strains, with low PROVEAN scores (L269F: − 3.8 and V268D: − 6.6). Previous work has shown that the reference rat strain precursor (BN/SsNHsd) has an abnormal pre-pulse inhibition (Palmer et al. 2003). This trait was linked to two regions in a QTL study (Palmer et al. 2003). The chromosome 2 region encompasses the *Dclk1* gene. This gene belongs to a family of three genes (along with *Dcx*, *Dclk2*) which are all three important for normal neuronal functions (Dijkmans et al. 2010). *Dckl1* has been proven to be important for proper neuronal development in mice, based on full KO phenotypes (Koizumi et al. 2006). The available data thus make the abnormal Dclk1 protein in BN/SsNHsd a good candidate to explain its pre-pulse inhibition trait.

Finally, the *Ratpost* tool also allows the direct comparison of any two strains. This comparison allows only a list of differing transcripts and their differences as it is not informative to provide a classification in this case. For example, *Tyr* gene can easily be identified by comparing protein-coding gene sequence differences between the albino strain SS/Jr and the pigmented strain BN/SsN.

In conclusion, a searchable online repository (which will be updated yearly) of variant alleles of protein-coding genes is now available at *Ratpost* and can lead to extensive exploration and exploitation of the naturally occurring mutant variants fixed in the 40 sequenced rat strains. Our analysis and database concerns protein-coding genes only, and an extension toward non-coding RNAs and a link towards mRNA expression levels might be considered in the future. Combined with the efficient CRISPR/Cas-based mutagenesis, the availability of these naturally occurring sequence variations in these 40 rat strains is an added value to couple sequence variations to phenotypes, and moreover, CRISPR/Cas could be considered to correct some unwanted variations in the reference strain, upgrading it to a higher standard reference.

Methods

Sequence variation data

We obtained the sequence variation data (SNPs, insertions and deletions) of the inbred strains from the ftp site of the rat genome database (Hermsen et al. 2015; Smith et al. 2020). Data were converted to bed (vcf2bed) and filtered with bedops (Neph et al. 2012) to retain only events overlapping with exons of coding genes, by using the element

Table 2 A list of variants that are unique to the reference strain with a PROVEAN score < − 2.5

Ensemble gene	Chr	Gene symbol	Ensemble transcript	AA variation	PROVEAN score
ENSRNOG00000026065	1	*Slit1*	ENSRNOT00000035415	Y869_D870insC	− 18.0
ENSRNOG00000026065	1	*Slit1*	ENSRNOT00000034758	Y869_D870insC	− 17.7
ENSRNOG00000032922	2	*Dclk1*	ENSRNOT00000078337	V268D	− 6.7
ENSRNOG00000032922	2	*Dclk1*	ENSRNOT00000093407	V268D	− 6.6
ENSRNOG00000032922	2	*Dclk1*	ENSRNOT00000093407	L269F	− 3.8
ENSRNOG00000032922	2	*Dclk1*	ENSRNOT00000078337	L269F	− 3.7
ENSRNOG00000030212	3	*Cerkl*	ENSRNOT00000073412	G313V	− 6.1
ENSRNOG00000030212	3	*Cerkl*	ENSRNOT00000043238	G362V	− 6.0
ENSRNOG00000030212	3	*Cerkl*	ENSRNOT00000079781	G269V	− 6.0
ENSRNOG00000030212	3	*Cerkl*	ENSRNOT00000079212	G341V	− 5.9
ENSRNOG00000036964	3	*Ralgapa2*	ENSRNOT00000015637	V429_K430insYR	− 21.3
ENSRNOG00000049125	3	*LOC684539*	ENSRNOT00000072178	C137S	− 4.8
ENSRNOG00000050251	3	*MGC105649*	ENSRNOT00000073569	D47N	− 4.6
ENSRNOG00000007603	5	*Gabrr1*	ENSRNOT00000010172	P430L	− 3.3
ENSRNOG00000023452	5	*AABR07050449.1*	ENSRNOT00000031978	S346F	− 7.9
ENSRNOG00000005965	7	*Irak4*	ENSRNOT00000007932	K408E	− 3.5
ENSRNOG00000032552	7	*Zfp799*	ENSRNOT00000037708	E385G	− 7.6
ENSRNOG00000032552	7	*Zfp799*	ENSRNOT00000087294	E346G	− 7.3
ENSRNOG00000000196	8	*Cyp19a1*	ENSRNOT00000000212	V156G	− 5.9
ENSRNOG00000003144	10	*Gprc5c*	ENSRNOT00000004256	P333Q	− 5.4
ENSRNOG00000003144	10	*Gprc5c*	ENSRNOT00000086924	P333Q	− 5.4
ENSRNOG00000003679	10	*Med13*	ENSRNOT00000035280	I848T	− 4.9
ENSRNOG00000007117	10	*Cluap1*	ENSRNOT00000064964	N362D	− 3.7
ENSRNOG00000007117	10	*Cluap1*	ENSRNOT00000078577	N373D	− 3.7
ENSRNOG00000000281	11	*Prodh1*	ENSRNOT00000082855	C312G	− 8.6
ENSRNOG00000000281	11	*Prodh1*	ENSRNOT00000002576	C373G	− 8.3
ENSRNOG00000049282	12	*Oas2*	ENSRNOT00000077542	G710W	− 12.7
ENSRNOG00000012850	15	*Kat6b*	ENSRNOT00000070893	P826A	− 2.7
ENSRNOG00000055935	15	*Olr1286*	ENSRNOT00000086118	I122R	− 7.2
ENSRNOG00000043387	16	*Cpe*	ENSRNOT00000064297	Y226E	− 6.6
ENSRNOG00000043387	16	*Cpe*	ENSRNOT00000064297	R232H	− 3.6
ENSRNOG00000052513	16	*Wapl*	ENSRNOT00000078390	Q574E	− 2.7
ENSRNOG00000059057	16	*Shld2*	ENSRNOT00000087737	N38K	− 3.3
ENSRNOG00000059057	16	*Shld2*	ENSRNOT00000085012	N38K	− 3.1
ENSRNOG00000002509	X	*Gnl3l*	ENSRNOT00000003397	N369K	− 4.8
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	M217_R218insYEAQQQ	− 31.1
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	P226_S227insADATRC	− 18.7
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	L211_S212insCT	− 16.8
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	I193D	− 7.4
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	W196_R198del	− 6.1
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	S222Q	− 3.3
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	Q194S	− 2.7
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	L199V	− 2.7
ENSRNOG00000003707	X	*Zmym3*	ENSRNOT00000076042	L195V	− 2.6
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000082902	G232V	− 5.8
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000078617	G273V	− 5.8
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000086952	G206V	− 5.7
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000082725	G54V	− 5.2
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000082902	G231V	− 4.8
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000078617	G272V	− 4.8

Table 2 (continued)

Ensemble gene	Chr	Gene symbol	Ensemble transcript	AA variation	PROVEAN score
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000086952	G205V	− 4.7
ENSRNOG00000057706	X	*Kdm5c*	ENSRNOT00000082725	G53V	− 4.5

of 1 option with the variant file as reference and exon locations as query input. All data from the RN5 version of the rat genome were converted to RN6 with the use of the picard tools (LiftoverVcf) (Institute 2019) using the RN5 to RN6 chain file obtained from UCSC.

Reference genome and annotation

We were constrained by the variation data to use the RN6 version of the reference genome. The sequences and corresponding annotation (gene transfer format; gtf) were downloaded from ensemble ftp server. The gtf file was converted to bed (gtf2bed) and filtered to only retain exons from transcripts containing at least one variant. For this, the filtered set of variants obtained as described in the above section was used as the query for bedops—element of 1, with the exon positions as the reference. Exon sequences were extracted from the reference genome fasta file with the getfasta command from bedtools using the filtered transcript set.

Transcript classification

We used an updated version of the script used in mousepost (Timmermans et al. 2017). The cDNA sequences were constructed using from the previously constructed exon sequences and split up in the 5′ UTR, CDS, and 3′ UTR. For each strain, the sequence variations present for that strain were applied to the CDS. The in silico mutation of the sequence was done using a custom made perl script from the 5′ end to the 3′ end of the sequences in order to easily take into account sequence offsets from indels. All genomic positions were converted to cDNA/CDS positions internally and applied in sequence, negative-strand sequences were reverse complemented for further processing. Finally, the CDS was translated to the protein sequence, using the standard vertebrate coding table. Classification was performed by comparing the reference and alternate aa sequences into three classes (SG, SL, MUT). For SG, any offsets from indels were taken into account, and SL variants were called on the loss of the canonical stop codon, in which case the 3′UTR, with variants applied (same as CDS), was appended to the CDS and a scan for a new stop codon was done. All variants not classed as SG/SL were placed as MUT, and all processing was stopped after encountering a SG-classed variant. Transcripts were assigned to the most severe variant class. We processed all strains in parallel.

BLAST databases

The BLAST+ program suite, v2.2.30, was pre-installed on the HPC infrastructure, and BLAST databases for the blast tools were obtained from the NCBI ftp site. The database with non-redundant protein sequences for use with PROVEAN was downloaded November 16, 2016 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). This will be used by the PROVEAN software, more specifically, the psi-blast program.

PROVEAN

Several programs have been developed to estimate the effect of a given sequence variation on the function of the protein. Because PROVEAN is able to interpret small insertions and deletions, this tool was selected (Choi and Chan 2015b). First, a file for each transcript in each strain was constructed with the mutated positions. For this, a global pairwise sequence alignment was constructed between the reference and alternative transcript with needle (EMBOSS tools) (Madeira et al. 2019). This alignment file, along with the positional data from the classification step, was processed with a perl script that was specifically created to build these files. To keep run times within acceptable limits, the PROVEAN tool was run on a high-performance computing (HPC) cluster for each strain in sequence; in this way, we could save and reuse the supported sequence sets. Due to the high computational requirements and limitations in HPC access, some variants were excluded from PROVEAN analysis: all variants from strain pairwise comparisons were excluded, as the PROVEAN scores are only truly relevant in comparison to a reference and variants from SL- and SG-classed transcripts were also not processed with PROVEAN. The score provided by PROVEAN depends in part on the number of available sequences for a given transcript. We have followed the suggestions of the authors of the PROVEAN tool and show a waring if the number of sequences is < 50, which may give unreliable results, and we use a score of -2.5 as a ceiling cut-off for potentially deleterious variants (Choi and Chan 2015b).

Gene ontology

The gene ontology (GO) annotation was downloaded from the gene ontology consortium website (www.geneontology.org). We processed this file to obtain the GO terms for all genes in our dataset to allow the GO search functionality in the web tool.

Data storage & presentation

All results from the analyses are stored in a MySQL relational database. Database tables are interlinked and have been optimized with indexes to allow fast searches. This database is made accessible through the website ratpost.be, built on the bootstrap 4 framework php and javascript. All tables found are created by use of the DataTables plugin.

Author contributions ST performed all the analysis and built the *Ratpost* website and wrote the manuscript. CL supervised the research and edited and finalized the manuscript.

Funding Research was funded by the Agency for Innovation of Science and Technology in Flanders (IWT), the Research Council of Ghent University (GOA program), the Research Foundation Flanders (FWO Vlaanderen), and the Interuniversity Attraction Poles Program of the Belgian Science Policy (IAP-VI-18).

Data availability All data used are publicly available at the rat genome database. All results can be found on the website mousepose.be.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval Not applicable.

Consent for publication Not applicable.

References

Afenjar A, Moutard ML, Doummar D, Guet A, Rabier D, Vermersch AI, Mignot C, Burglen L, Heron D, Thioulouse E, de Villemeur TB, Campion D, Rodriguez D (2007) Early neurological phenotype in 4 children with biallelic PRODH mutations. Brain Dev 29:547–552

Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, Tschannen MR, Kaisaki PJ, Otto GW, Ma MC, Keane TM, Hummel O, Saar K, Chen W, Guryev V, Gopalakrishnan K, Garrett MR, Joe B, Citterio L, Bianchi G, McBride M, Dominiczak A, Adams DJ, Serikawa T, Flicek P, Cuppen E, Hubner N, Petretto E, Gauguier D, Kwitek A, Jacob H, Aitman TJ (2013) Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. Cell 154:691–703

Barefield DY, Puckelwartz MJ, Kim EY, Wilsbacher LD, Vo AH, Waters EA, Earley JU, Hadhazy M, Dellefave-Castillo L, Pesce LL, McNally EM (2017) Experimental modeling supports a role for MyBP-HL as a novel myofilament component in arrhythmia and dilated cardiomyopathy. Circulation 136:1477–1491

Bergman I, Basse PH, Barmada MA, Griffin JA, Cheung NK (2000) Comparison of in vitro antibody-targeted cytotoxicity using mouse, rat and human effectors. Cancer Immunol Immunother CII 49:259–266

Blaszczyk WM, Arning L, Hoffmann KP, Epplen JT (2005) A Tyrosinase missense mutation causes albinism in the Wistar rat. Pigment Cell Res 18:144–145

Choi Y, Chan AP (2015a) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics (Oxford, England) 31:2745–2747

Choi Y, Chan AP (2015b) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7:e46688

Conrad CH, Brooks WW, Hayes JA, Sen S, Robinson KG, Bing OH (1995) Myocardial fibrosis and stiffness with hypertrophy and heart failure in the spontaneously hypertensive rat. Circulation 91:161–170

Dijkmans TF, van Hooijdonk LW, Fitzsimons CP, Vreugdenhil E (2010) The doublecortin gene family and disorders of neuronal structure. Cent Nerv Syst Agents Med Chem 10:32–46

Ellenbroek B, Youn J (2016) Rodent models in neuroscience research: is it a rat race? Dis Models Mech 9:1079–1087

EUR-Lex—52020DC0016—EN—EUR-Lex

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle

SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F, Rat Genome Sequencing Project C (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

Hedrich HJ (2000) History, strains and models. In: Krinke GJ (ed) The laboratory rat. Elsevier, pp 3–16

Hermsen R, de Ligt J, Spee W, Blokzijl F, Schafer S, Adami E, Boymans S, Flink S, van Boxtel R, van der Weide RH, Aitman T, Hubner N, Simonis M, Tabakoff B, Guryev V, Cuppen E (2015) Genomic landscape of rat strain and substrain variation. BMC Genom 16:357

Holl K, He H, Wedemeyer M, Clopton L, Wert S, Meckes JK, Cheng R, Kastner A, Palmer AA, Redei EE, Solberg Woods LC (2018) Heterogeneous stock rats: a model to study the genetics of despair-like behavior in adolescence. Genes Brain Behav 17:139–148

Institute B (2019) Picard toolkit. In: Broad Institute, GitHub repository

Jonckers E, Van Audekerke J, De Visscher G, Van der Linden A, Verhoye M (2011) Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse. PLoS One 6:e18876

Koizumi H, Tanaka T, Gleeson JG (2006) Doublecortin-like kinase functions with doublecortin to mediate fiber tract decussation and neuronal migration. Neuron 49:55–66

Leong XF, Ng CY, Jaarin K (2015) Animal models in cardiovascular research: hypertension and atherosclerosis. BioMed Res Int 2015:528757

Liu T, Li H, Ding G, Wang Z, Chen Y, Liu L, Li Y, Li Y (2015) Comparative genome of GK and Wistar rats reveals genetic basis of type 2 diabetes. PLoS One 10:e0141859

Logan CA (1999) The altered rationale for the choice of a standard animal in experimental psychology: Henry H. Donaldson, Adolf Meyer, and "the" albino rat. Hist Psychol 2:3–24

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641

Modlinska K, Pisula W (2020) The Norway rat, from an obnoxious pest to a laboratory pet. eLife 9:e50651

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, Sandstrom R, Humbert R, Stamatoyannopoulos JA (2012) BEDOPS: high-performance genomic feature operations. Bioinformatics 28:1919–1920

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-745

Obrosova IG, Drel VR, Kumagai AK, Szabo C, Pacher P, Stevens MJ (2006) Early diabetes-induced biochemical changes in the retina: comparison of rat and mouse models. Diabetologia 49:2525–2533

Palmer AA, Breen LL, Flodman P, Conti LH, Spence MA, Printz MP (2003) Identification of quantitative trait loci for prepulse inhibition in rats. Psychopharmacology 165:270–279

Perry TL, Hardwick DF, Lowry RB, Hansen S (1968) Hyperprolinaemia in two successive generations of a North American Indian family. Ann Hum Genet 31:401–407

Philipeaux JM (1856) Note sur l'exstirpation des capsules surrenales chez les rats albinos (Mus rattus). Comptes rendus hebdomadaires des séances de l'Académie des sciences 43:904–906

Podcheko A, Northcott P, Bikopoulos G, Lee A, Bommareddi SR, Kushner JA, Farhang-Fallah J, Rozakis-Adcock M (2007) Identification of a WD40 repeat-containing isoform of PHIP as a novel regulator of beta-cell growth and survival. Mol Cell Biol 27:6484–6496

Richter CP (1959) Rats, man, and the welfare-state. Am Psychol 14:18–28

Savory W (1863) Experiments on food; its destination and uses. Lancet 81:381–383

Sharp PE, Villano JS (2013) The laboratory rat, 2nd edn. CRC Press, Boca Raton

Smith JR, Hayman GT, Wang SJ, Laulederkind SJF, Hoffman MJ, Kaldunski ML, Tutaj M, Thota J, Nalabolu HS, Ellanki SLR, Tutaj MA, De Pons JL, Kwitek AE, Dwinell MR, Shimoyama ME (2020) The year of the rat: the rat genome database at 20: a multi-species knowledgebase and analysis platform. Nucleic Acids Res 48:D731–D742

Stumvoll M, Goldstein BJ, van Haeften TW (2005) Type 2 diabetes: principles of pathogenesis and therapy. Lancet (London, England) 365:1333–1346

Timmermans S, Libert C (2018) Overview of inactivating mutations in the protein-coding genome of the mouse reference strain C57BL/6J. JCI Insight 3(13):e121758

Timmermans S, Van Montagu M, Libert C (2017) Complete overview of protein-inactivating sequence variations in 36 sequenced mouse inbred strains. Proc Natl Acad Sci USA 114:9158–9163

Trippodo NC, Frohlich ED (1981) Similarities of genetic (spontaneous) hypertension. Man and rat. Circ Res 48:309–319

Watson JB (1914) Behavior: an introduction to comparative psychology. Holt, Rinehart and Winston, New York

Woods LC, Mott R (2017) Heterogeneous stock populations for analysis of complex traits. Methods Mol Biol (Clifton, NJ) 1488:31–44