Pooled analysis of radiation hybrids identifies loci for growth and drug action in mammalian cells: Supplemental Material

Arshad H. Khan[1], Andy Lin[2], Richard T. Wang[3], Joshua S. Bloom[3,4], Kenneth Lange[3], Desmond J. Smith[1]
[1] Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA, Box 951735, Los Angeles, CA 90095-1735, USA
[2] Office of Information Technology, UCLA, Box 951557, 5308 MS, Los Angeles, CA 90095-1557, USA
[3] Department of Human Genetics, David Geffen School of Medicine, UCLA, Box 957088, Los Angeles, CA 90095-7088, USA
[4] Howard Hughes Medical Institute, David Geffen School of Medicine, UCLA, Box 957088, Los Angeles, CA 90095-7088, USA
Corresponding author: DSmith@mednet.ucla.edu

1 Cells

Human embryonic kidney 293 (HEK293) cells were obtained from the American Type Culture Collection (Manassas, VA) around 2006. These cells are female (Lin et al. 2014). We validated the HEK293 cells in a previous study using simple tandem repeat genotyping (Khan et al. 2016).

Additional confirmation was obtained from the low pass genome sequence data generated during this study. A total of 1366 known common homozygous coding region single nucleotide polymorphisms (SNPs) had previously been identified in the HEK293 genome (Lin et al. 2014). Of these SNPs, a total of 640 were identified by our low pass sequencing and 637 (99.5 %) agreed with the known allelic variants ($\chi^2[640] = 982, P < 2.2 \times 10^{-16}$, likelihood ratio test).

Similarly, 107 novel homozygous coding region SNPs had been previously identified in the HEK293 genome (Lin et al. 2014), of which 57 were identified in our low pass data. A total of 32 of these 57 SNPs (56 %) agreed with the previously identified alleles ($\chi^2[57] = 89, P = 4.5 \times 10^{-3}$, likelihood ratio test).

Further, there was high concordance between the copy number alterations (CNAs) in our HEK293 isolate and the published HEK293 genome sequence (Lin et al. 2014) ($R = 0.52$, $t[1{,}28173] = 101.1$, $P < 2.2 \times 10^{-16}$) (Fig. 2C; Supplemental Fig. S4).

A23 hamster cells were kindly donated by Dr Christine Farr (University of Cambridge, Cambridge, UK), around 2004. These cells are a Chinese hamster lung fibroblast-derived cell line, a descendant of the DON cell line and are male (Park et al. 2008). We had previously confirmed the A23 cells were of Chinese hamster origin by simple tandem repeat genotyping (Khan et al. 2016). The hamster origin of the cells was further verified by the stringent alignment of our low pass sequence data to the published hamster genome (Rupp et al. 2018).

HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) with 10 % fetal bovine serum (FBS) and 1 × penicillin/streptomycin (100 units ml^{-1} penicillin, 100 μg ml^{-1} streptomycin; Thermo Fisher Scientific®). The A23 cells were grown in α minimum essential medium (MEM), supplemented with 10 % fetal bovine serum (FBS), 1 × penicillin/streptomycin. The RH pools were propagated in the same medium as the A23 cells, with the addition of 1 × HAT (100 μM hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine; Thermo Fisher Scientific®) (Park et al. 2008; Wang et al. 2011).

Cells were evaluated for mycoplasma contamination by aligning sequence data to the relevant bacterial genomes.

2 Paclitaxel

Paclitaxel (Taxol®) was obtained from Tocris®. Dimethyl sulfoxide (DMSO) was from Thermo Fisher Scientific®. To aid solubilization, paclitaxel was dissolved in DMSO and then diluted in medium. The final concentration of DMSO in all samples was 0.37 %.

3 Cell fusion

We performed cell fusions as described (Schafer, Farr 1998), with modifications. Four separate fusion reactions of equal size were used to constitute each pool, with irradiation employing either 30 Gray (Gy) or 100 Gy from a Shepherd Mark I $^{137}_{55}$Cs Irradiator (Supplemental Table S1). We explored possible cost savings in plasticware and culture medium by seeding the reactions into HAT selective medium using both a 1:5 dilution (five 75 cm^2 tissue flasks) as well as the 1:10 dilution (ten flasks) recommended in published protocols (Schafer, Farr 1998). The radiation dose and dilution conditions were applied in balanced fashion.

The expected fragment size for the two radiation doses was 10 Mb and 4 Mb, respectively (Cox et al. 1990; Hudson et al. 2001; Kwitek et al. 2004; McCarthy 1996; McCarthy et al. 2000; Olivier et al. 2001;

Walter et al. 1994). The mean fragment size weighted by clone number at each dose was 7.4 Mb ± 0.1 Mb for each pool. To evaluate the frequency of spontaneous revertants, mock fused A23 cells were also plated.

When the four fusion reactions for each of the six pools reached confluence, the cells were consolidated and a small portion reserved for genotyping. These samples constitute week 0 and 0 nM paclitaxel and are referred to as the "RH pools" (Supplemental Table S2). The remainder of the pools were grown in HAT medium for 1, 2, 3, 4 or 6 weeks, supplemented with 0, 8, 25 or 75 nM paclitaxel, and analyzed using low pass sequencing.

4 Clone, colony and revertant counts

4.1 RH clone counts

The RH clone counts were significantly overdispersed. The dispersion was estimated using the ratio of the sum of the squared Pearson residuals to the residual degrees of freedom (dispersion = 25, $\chi^2[23] = 570$, $P < 2.2 \times 10^{-16}$, Poisson model). Since a dispersion value of 1 is expected for the Poisson model, the significant overdispersion suggests a poor fit. In addition, there was a potential batch effect of RH pool (Supplemental Figs. S1A, S11; Supplemental Table S1). To account for the overdispersion and batch effects, a negative binomial mixed model was used to evaluate the RH clone counts

$$log(C_{ijk}) \sim \beta_0 + \beta_1 D_i + \beta_2 R_j + \beta_3 D_i R_j + P_k + \epsilon_{ijk}, \quad (1)$$

where C_{ijk} is the number of clones, β_0 is the fixed effect of intercept, β_1 is the fixed effect of dilution D, at $i = 5$ and 10-fold, β_2 is the fixed effect of radiation dose R, at $j = 30$ and 100 Gy, β_3 is the fixed effect of the interaction $D_i R_j$, P is the random intercept due to RH pool with $k = 1$–6 and ϵ_{ijk} is the residual variance. We used the same model to investigate the colony and revertant counts.

The model was analyzed using R 3.5.0 (R Core Team 2018) and glmmTMB (Brooks et al. 2017) employing restricted maximum likelihood (REML) estimation. Post hoc analyses to obtain P values employed emmeans (Lenth 2019). Average conditional effects of dilution were evaluated at a radiation dose of 65 Gy, and average conditional effects of radiation at a dilution of 7.5-fold.

The dispersion parameter estimated from the negative binomial mixed model was not significantly different from unity (dispersion = 0.70, $\chi^2[22] = 15$, $P = 0.85$), indicating a superior fit to the Poisson model. Consistent with a batch effect of RH pool, the random intercept also significantly improved model fit (Akaike information criterion, AIC = 387 without random effect, 378 with random effect; $\chi^2[1] = 11$, $P = 9.5 \times 10^{-4}$, likelihood ratio test). The adjusted between- and within-pool variance accounted for 69 % (intraclass correlation coefficient, ICC = 0.69) and 31 % of the total variance components, respectively.

The higher radiation dose resulted in a significantly smaller number of RH clones ($t[1,22] = 2.4, P = 0.02$; average conditional effect) (Supplemental Fig. S1A). The 1:10 dilution recommended in the literature (Schafer, Farr 1998) was superior to the 1:5 dilution, providing a highly significant 2.6 ± 0.3-fold increase in the number of RH clones ($t[1,22] = 8.5, P = 2.3 \times 10^{-8}$; average conditional effect) (Supplemental Fig. S1A). The interaction of radiation dose and dilution was non-significant ($t[1,22] = 0.69, P = 0.50$).

4.2 The RH clone, colony and revertant counts are correlated

The number of RH clones was highly correlated with both the total number of colonies ($R = 0.99$, $t[1,22] = 34$, $P < 2.2 \times 10^{-16}$) and the number of revertants ($R = 0.85$, $t[1,22] = 7.5$, $P = 1.7 \times 10^{-7}$). However, there were some differences in the effects of radiation dose and dilution on the colonies and revertants compared to the RH clones.

4.3 Colony counts

Similar to the RH clones, the colonies showed significant overdispersion (dispersion = 20, $\chi^2[23] = 5 \times 10^2$, $P < 2.2 \times 10^{-16}$, Poisson model). In contrast, the dispersion parameter estimated from the negative binomial mixed model was not significantly different from unity (dispersion = 0.72, $\chi^2[22] = 16$, $P = 0.82$) (Supplemental Table S1). Further, the random effect of RH pool significantly improved model fit (AIC = 400 without random effects, 384 with random effects; $\chi^2[1] = 18$, $P = 2.0 \times 10^{-5}$, likelihood ratio test, ICC = 0.82) (Supplemental Figs. S1B, S11).

Like the RH clones, the higher radiation dose significantly decreased colony numbers (t[1,22] = 2.8, $P = 0.01$; average conditional effect) (Supplemental Fig. S1B). However, the dominant effect was again dilution, with the higher dilution giving larger colony counts (t[1,22] = 11, $P = 1.9 \times 10^{-10}$; average conditional effect) (Supplemental Fig. S1B). The interaction of dilution and radiation dose was insignificant (t[1,22] = 0.06, $P = 0.96$).

4.4 Revertant counts

There was also significant overdispersion for the revertants (dispersion = 3.5, $\chi^2[23] = 80$, $P = 2.7 \times 10^{-8}$, Poisson model). Further, the dispersion parameter estimated from the negative binomial mixed model was not significantly different from unity (dispersion = 0.72, $\chi^2[22] = 16$, $P = 0.82$) (Supplemental Table S1). The random effect of RH pool significantly improved model fit (AIC = 339 without random effects, 318 with random effects; $\chi^2[1] = 23$, $P = 1.9 \times 10^{-6}$, likelihood ratio test, ICC = 0.86) (Supplemental Figs. S1C, S11).

There were no significant fixed effects of radiation for the revertants (t[1,22] = 1.5, $P = 0.15$; average conditional effect) (Supplemental Fig. S1C). The most significant effect was dilution, with higher dilution giving larger numbers (t[1,22] = 11, $P = 3.5 \times 10^{-10}$; average conditional effect) (Supplemental Fig. S1C). The revertants showed a significant interaction of dilution and radiation dose (t[1,22] = 2.2, $P = 0.04$).

5 Low pass sequencing

Genotyping employed Illumina sequencing with single-end reads of 64 bp. As background information, we obtained 41.7 M sequence reads from the HEK293 donor cells and 37.2 M reads from the A23 recipient cells (Supplemental Tables S3, S4). In addition, we obtained 41.1 ± 2.8 M reads for each of the six RH pools (Supplemental Table S5). For all RH samples under the various conditions of growth and drug, including the six RH pools, we obtained 40.5 ± 0.7 M reads per sample, corresponding to a sequencing depth of 0.84 ± 0.01 times the human genome.

5.1 Human and hamster genomes

The human genome is a high quality finished assembly. The hamster genome is a high quality draft assembly with a scaffold N50 of 20.2 Mb, L50 of 32 , 0.12 % gaps and 96.6 % of the sequence assigned to a chromosome (Rupp et al. 2018). In fact, the continuity of the hamster genome is superior to the rat and approaches that of the mouse. However, using selected human/hamster sequence variants would neglect genome regions that could not be aligned between the two species. We therefore decided to maximize information recovery by aligning all sequence reads. Stringent alignment parameters were employed, so that only a small proportion of reads mapped across species (Supplemental Tables S3, S4). Using the HEK293 and A23 sequence data, the misalignment rate for the human to hamster nuclear genomes was 0.13 % and 0.27 % in the other direction. For the mitochondrial genomes, the human to hamster misalignment rate was 0.19 % and 0.14 % reversed.

5.2 Alignment strategies for human DNA

Although only 0.27 % of hamster reads also aligned to the human nuclear genome (Supplemental Table S3), this low frequency became non-negligible in the context of the six RH pools, where there was a mean of 59 ± 14-fold more mapped hamster reads than human (Supplemental Tables S5, S6). To evaluate the human DNA in the RH pools with greater specificity, we discarded reads that aligned to both species (Fig. 1). A mean of 0.88 ± 0.11 M reads initially aligned to the human genome (Supplemental Table S5). After discarding 0.13 ± 0.03 M reads that showed cross-species alignment (a hamster-to-human misalignment rate of $16.8\,\% \pm 3.7\,\%$), 0.74 ± 0.12 M reads specifically mapped to the human genome in the six RH pools. The corresponding density of human reads was 0.015 ± 0.0024 times the human genome.

For all samples under the various conditions of growth time and drug exposure, including the six RH pools, there was a mean of 84 ± 16-fold more mapped hamster reads than human. A mean of 0.90 ± 0.04 M reads initially aligned to the human genome. Of these, 0.16 ± 0.005 M reads ($21.0\,\% \pm 1.1\,\%$) also aligned to the hamster genome and were discarded, giving 0.75 ± 0.04 M reads specifically aligned to the human genome. The corresponding density of human reads was 0.015 ± 0.0008 times the human genome.

5.3 Alignment strategies for hamster DNA

A total of 0.13 % of human reads from HEK293 cells also mapped to hamster (Supplemental Table S3). However, the much smaller amount of human DNA in the RH samples meant that the impact of cross-species alignment was much less for the hamster genome than the human. Nevertheless, to accurately quantitate the levels of hamster DNA, we discarded hamster reads that cross-aligned to both species (Fig. 1).

A mean of 36.4 ± 2.5 M reads initially aligned to the hamster genome in the six RH pools, giving 36.2 ± 2.5 M reads specific to hamster after discarding 0.13 ± 0.03 M reads that aligned to both hamster and human (a cross-alignment rate of $0.36\,\% \pm 0.06\,\%$) (Supplemental Table S6). The corresponding read density was 0.98 times the hamster genome $\pm\, 0.07$.

For all samples, a mean of 35.8 ± 0.6 M reads initially aligned to the hamster genome. Of these, 0.16 ± 0.005 M reads ($0.45\,\% \pm 0.01\,\%$) also aligned to the human genome and were discarded, giving 35.6 ± 0.6 M reads specifically aligned to the hamster genome. The corresponding density of hamster reads was 0.96 ± 0.02 times the hamster genome.

5.4 Alignment strategies for mitochondrial genomes

A total of 0.19 % of human reads cross-aligned with hamster mitochondria and 0.14 % of hamster reads cross-aligned with human mitochondria (Supplemental Table S4). After correcting for cross-species alignments, 172 ± 99 reads mapped to the human mitochondrial genome in the six RH pools (Supplemental Table S12). For all samples under the various conditions of growth time and drug concentration, including the six RH pools, 68 ± 15 reads aligned to the human mitochondrial genome after removal of hamster cross-alignments. The low numbers of human mitochondrial sequence reads suggest that these organelles were transferred with low efficiency to the hamster cells.

After correcting for cross-species alignments, $92\,926 \pm 6797$ reads mapped to the hamster mitochondrial genome for each of the six RH pools (Supplemental Table S12). For all samples, including the six RH pools, $90\,434 \pm 3033$ reads aligned to the hamster mitochondrial genome after removal of human cross-alignments. All analyses of human and hamster nuclear and mitochondrial DNA used read counts purged of cross-species alignments, unless otherwise noted.

6 DNA retention in the RH pools

6.1 The hamster genome is unchanged in RH cells

The copy number of the hamster genome for each 1 Mb window in the A23 and RH cells was calculated as $C^{\mathrm{G}}_{\mathrm{Ha}} = N^{\mathrm{G}}_{\mathrm{Ha}}/\bar{N}^{\mathrm{G}}_{\mathrm{Ha}}$, where $N^{\mathrm{G}}_{\mathrm{Ha}}$ is the number of hamster-specific reads in each window and $\bar{N}^{\mathrm{G}}_{\mathrm{Ha}}$ is the mean number of hamster-specific reads across all windows.

There was strong similarity between copy number alterations (CNAs) in the parental A23 cells and the RH pools, suggesting that the hamster genome in the six RH pools was not substantially altered by the cell fusions ($R = 0.98$, $t[1{,}2041] = 253$, $P < 2.2 \times 10^{-16}$, non-overlapping 1 Mb windows) (Fig. 2A,B; Supplemental Figs. S2, S3A).

6.2 Regions of enhanced human DNA retention in the RH cells

The copy number of the human genome for each 1 Mb window in the HEK293 cells was calculated as $C^{\mathrm{G}}_{\mathrm{Hu}} = N^{\mathrm{G}}_{\mathrm{Hu}}/\bar{N}^{\mathrm{G}}_{\mathrm{Hu}}$, where $N^{\mathrm{G}}_{\mathrm{Hu}}$ is the number of reads in each window and $\bar{N}^{\mathrm{G}}_{\mathrm{Hu}}$ is the mean number of reads across all windows.

While the recipient hamster genome was largely unchanged by the cell fusion, human DNA donated to the RH cells was retained as a selection of extra DNA fragments (Fig. 2C,D; Supplemental Fig. S5). The thymidine kinase (TK1) gene was used as a selectable marker and therefore has a retention frequency of 1 (Fig. 2F). The retention of the human genome in the RH pools relative to *TK1* was calculated as $R^{TK1}_{\mathrm{Hu}} = N^{\mathrm{G}}_{\mathrm{Hu}}/N^{TK1}_{\mathrm{Hu}}$, where $N^{\mathrm{G}}_{\mathrm{Hu}}$ is the number of human-specific reads in each window and N^{TK1}_{Hu} is the maximum number of human-specific reads in a 1 Mb window encompassing *TK1*.

Since the probability of human gene transfer from the donor cells increases with copy number, there was a significant correlation between human DNA retention in the six RH pools and copy number in the HEK293 cells ($R = 0.47$, $t[1{,}3007] = 29$, $P < 2.2 \times 10^{-16}$, non-overlapping 1 Mb windows) (Supplemental Fig. S3B).

Donor centromeres and telomeres confer increased stability on the human DNA fragments in RH cells (Wang et al. 2011). In fact, nearly all centromeres showed significantly increased retention in the RH pools (centromere to non-centromere read ratio = 1.1 in HEK293 cells; 5.5 ± 0.2 in the six RH pools; $t[1{,}5] = 7.4$, $P = 7.0 \times 10^{-4}$, Two Sample t-test with pooled variance) (Fig. 2C,D; Supplemental Fig. S5).

The combination of pooling and low-pass sequencing employed in this study more clearly demonstrated augmented centromere retention in the RH cells than older technologies such as arrays or PCR, which had few centromeric markers. For Chromosome 6, the centromere copy number was greater than 1, indicating highly efficient retention and replication (Fig. 2D; Supplemental Fig. S5). However, centromere sequences in the GRCh38/hg38 human genome assembly represent an average of sets of chromosome-specific α-satellite repeats (Contreras-Galindo et al. 2017; Schneider et al. 2017). Caution is therefore warranted when interpreting centromere retention values.

6.3 Human DNA retention frequency

Non-centromeric regions showed relatively constant retention across the genome, as previously documented in RH panels (Avner et al. 2001; Hudson et al. 2001; Kwitek et al. 2004; McCarthy et al. 2000; Olivier et al. 2001) (Fig. 2D; Supplemental Fig. S5). We calculated the overall retention of the human genome in the RH pools in two ways (Supplemental Table S7). The first compared the median number of human-specific sequence reads in non-overlapping 1 Mb windows to the median number of hamster-specific reads. A multiplier of two was applied since the retained human fragments are overwhelmingly haploid, while the hamster genome is diploid. The median retention frequency was calculated as $\tilde{R}^{\mathrm{G}}_{\mathrm{Hu}} = 2(\tilde{N}^{\mathrm{G}}_{\mathrm{Hu}}/\tilde{N}^{\mathrm{G}}_{\mathrm{Ha}})$, where $\tilde{N}^{\mathrm{G}}_{\mathrm{Hu}}$ is the median number of human-specific reads in non-overlapping 1 Mb genome windows and $\tilde{N}^{\mathrm{G}}_{\mathrm{Ha}}$ is the median number of hamster-specific reads.

The second approach measured the median retention of the human genome relative to *TK1*, with a correction factor to account for the revertants. Thus, $\tilde{R}_{\mathrm{Hu}}^{TK1} = (\tilde{N}_{\mathrm{Hu}}^{\mathrm{G}}/N_{\mathrm{Hu}}^{TK1})(1-r)$, where $\tilde{N}_{\mathrm{Hu}}^{\mathrm{G}}$ is the median number of human-specific reads in non-overlapping 1 Mb genome windows, N_{Hu}^{TK1} is the maximum number of human-specific reads in a 1 Mb window encompassing *TK1*, and r is the reversion frequency (Supplemental Table S1). Median values were used in calculating overall retention frequencies to avoid skewing of the results by the outlier values of the centromeres and *TK1*.

Using the first method, the mean of the median fraction of the human genome that was retained in the six RH pools was $2.6\,\% \pm 0.5\,\%$ and using the second $5.3\,\% \pm 0.3\,\%$ ($t[1{,}8.7] = 4.5$, $P = 1.6 \times 10^{-3}$, Welch's Two Sample t-test) (Supplemental Table S7). The mean retention frequency using the two methods was $3.9\,\% \pm 0.3\,\%$.

The lowest euchromatic retention frequency was at 54 530 000 bp on Chromosome 19 and was significantly greater than zero ($0.4\,\% \pm 0.1\,\%$, $t[1{,}5] = 4.2$, $P = 8.2 \times 10^{-3}$ using sequence alignments; $0.9\,\% \pm 0.1\,\%$, $t[1{,}5] = 7.8$, $P = 5.7 \times 10^{-4}$ using *TK1*; $0.7\,\% \pm 0.1\,\%$, $t[1{,}5] = 6.7$, $P = 1.1 \times 10^{-3}$, using mean of sequence alignments and *TK1*; One Sample t-tests).

7 Further details on copy number changes in the RH samples

After 1, 2, 3, 4 and 6 weeks of growth in HAT medium supplemented with 0, 8, 25 or 75 nM paclitaxel, the RH samples were analyzed by low pass sequencing. There were a total of 115 samples, including the six RH pools (Supplemental Table S2). We obtained data for all permutations of time and drug, except one, corresponding to $91.3\,\% \pm 5.3\,\%$ of all possible samples for each table cell.

7.1 Changes in human DNA retention

A linear mixed model was used to evaluate changes in overall human DNA retention in the RH samples

$$\tilde{R}_{\mathrm{Hu}_{ijkl}}^{\mathrm{G\ or\ }TK1} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \tag{2}$$

where $\tilde{R}_{\mathrm{Hu}_{ijkl}}^{\mathrm{G\ or\ }TK1}$ is the median human DNA retention frequency evaluated either by comparison to hamster-specific reads or to human *TK1* reads respectively (Section 6.3), β_0 is the fixed effect of intercept, β_1 is the fixed effect of growth G at time $i = 1$, 2, 3, 4 and 6 weeks, β_2 is the fixed effect of drug D at concentration $j = 0$, 8, 25 and 75 nM paclitaxel, β_3 is the fixed effect of the growth and drug interaction $G_i D_j$, P is the random intercept due to RH pool $k = 1$–6, L is a random intercept nested inside P_k to account for possible differences in drug histories of the RH pools, where $l = j \cdot n$ with $n = 0$ for week 0 and 1 for other times, and ϵ_{ijkl} is the residual variance. The addition of L_l was precautionary, since the samples were well balanced with respect to drug history.

The mixed model was analyzed using the lmer function of the lme4 package without an offset and employing restricted maximum likelihood (REML) estimation (Bates et al. 2015). Post hoc analyses to obtain P values employed emmeans (Lenth 2019). Average conditional effects of growth were evaluated at 27 nM paclitaxel and average conditional effects of paclitaxel at 3.2 weeks of growth.

There were significant decreases in retention as a result of growth ($2.6\,\% \pm 0.5\,\%$ at week 0 to $1.0\,\% \pm 0.2\,\%$ at week 6, $t[1{,}91.2] = 8.8, P = 6.8 \times 10^{-14}$, using sequence alignments; $5.3\,\% \pm 0.3\,\%$ at week 0 to $1.3\,\% \pm 0.2\,\%$ at week 6, $t[1{,}91.5] = 15, P < 2.2 \times 10^{-16}$, using *TK1*; $3.9\,\% \pm 0.3\,\%$ at week 0 to $1.1\,\% \pm 0.2\,\%$ at week 6, $t[1{,}90.8] = 15, P < 2.2 \times 10^{-16}$, using mean of sequence alignments and *TK1*; Kenward-Roger degrees of freedom, df; average conditional effects), but not paclitaxel ($t[1{,}24.5] = 1.7, P = 0.1$, sequence alignments; $t[1{,}24.7] = 0.07, P = 0.9$, *TK1*; $t[1{,}24.4] = 0.8, P = 0.5$, mean; Kenward-Roger df; average conditional effects) (Supplemental Figs. S7–S9). The interaction was also insignificant ($t[1{,}93.2] = 0.8, P = 0.5$, sequence alignments; $t[1{,}93.6] = 0.3, P = 0.7$, *TK1*; $t[1{,}92.8] = 0.7, P = 0.5$, mean; Kenward-Roger df).

There was inconsistent significance of the random effect of pool ($\chi^2[1] = 13, P = 4.1 \times 10^{-4}$, ICC = 0.49, sequence alignments; $\chi^2[1] = 4.4, P = 0.04$, ICC = 0.26, *TK1*; $\chi^2[1] = 1.8, P = 0.17$, ICC = 0.18, mean; likelihood ratio tests) and also drug history ($\chi^2[1] = 2.9, P = 0.09$, ICC = 0.18, sequence alignments; $\chi^2[1] = 4.9, P = 0.03$, ICC = 0.23, *TK1*; $\chi^2[1] = 4.7, P = 0.03$, ICC = 0.33, mean; likelihood ratio tests).

7.2 Changes in human retention and hamster copy number across genome

Copy number changes across the hamster genome in the RH pools were calculated as changes in the $\log_2$ transformed quantity $C^{\mathrm{G}}_{\mathrm{Ha}}$ (Section 6.1). Copy number changes across the human genome were calculated as changes in the $\log_2$ transformed retention $R^{\mathrm{G}}_{\mathrm{Hu}} = 2(N^{\mathrm{G}}_{\mathrm{Hu}}/\bar{N}^{\mathrm{G}}_{\mathrm{Ha}})$, where $N^{\mathrm{G}}_{\mathrm{Hu}}$ is the number of human-specific reads in each window and $\tilde{N}^{\mathrm{G}}_{\mathrm{Ha}}$ is the median number of hamster-specific reads (c.f. Section 6.3).

The hamster genome was relatively stable across the genome (Fig. 3A–D; Supplemental Fig. S10). In contrast, the human genome showed significant decreases in average copy number and increases in variance as a result of growth or paclitaxel, consistent with selective pressure.

Congruent with the overall decrease in human DNA retention in the RH samples (Supplemental Figs. S7–S9), we found significant genome-wide decreases in relative normalized human copy number at week 4 due to both growth (mean $\log_2$ copy number change = -0.54 ± 0.01, t[1,2911] = 48, $P < 2.2 \times 10^{-16}$, One Sample t-test, 4 weeks compared to 0 weeks, 0 nM paclitaxel, non-overlapping 1 Mb windows) (Fig. 3B) and also paclitaxel (mean $\log_2$ copy number change = -2.14 ± 0.02, t[1,2723] = 87, $P < 2.2 \times 10^{-16}$, One Sample t-test, 75 nM compared to 0 nM paclitaxel, 4 weeks, non-overlapping 1 Mb windows) (Fig. 3D).

Significant genome-wide decreases in relative normalized human copy number were also found at week 6 for both growth (mean $\log_2$ copy number change = -0.78 ± 0.01, t[1,2911] = 54.1, $P < 2.2 \times 10^{-16}$, One Sample t-test, 6 weeks compared to 0 weeks, 0 nM paclitaxel, non-overlapping 1 Mb windows) (Supplemental Fig. S10B) and paclitaxel (mean $\log_2$ copy number change = -1.46 ± 0.02, t[1,2781] = 64, $P < 2.2 \times 10^{-16}$, One Sample t-test, 75 nM compared to 0 nM paclitaxel, 6 weeks, non-overlapping 1 Mb windows) (Supplemental Fig. S10D).

The variance of relative normalized copy number changes for human DNA was significantly greater than hamster for both growth and paclitaxel at 4 weeks. The standard deviation (s.d.) of $\log_2$ copy number changes for 4 weeks of growth compared to 0 weeks (0 nM paclitaxel) was 0.05 ($\pm 4.1 \times 10^{-7}$) for the hamster genome and 0.60 ($\pm 2.6 \times 10^{-6}$) for the human (F[1,4953] = 3277, $P < 2.2 \times 10^{-16}$, Levene's test, non-overlapping 1 Mb windows) (Fig. 3A,B). Similarly, the s.d. of $\log_2$ copy number changes at 75 nM compared to 0 nM paclitaxel (4 weeks) was 0.10 ($\pm 7.6 \times 10^{-7}$) for the hamster genome and 1.3 ($\pm 6.1 \times 10^{-6}$) for the human (F[1,4765] = 2754, $P < 2.2 \times 10^{-16}$, Levene's test, non-overlapping 1 Mb windows) (Fig. 3C,D).

The variance of relative normalized copy number changes for human DNA was also significantly greater than hamster at 6 weeks. The s.d. of $\log_2$ copy number changes for 6 weeks of growth compared to 0 weeks (0 nM paclitaxel) was 0.07 ($\pm 5.7 \times 10^{-7}$) for the hamster genome and 0.78 ($\pm 3.3 \times 10^{-6}$) for the human (F[1,4953] = 3079, $P < 2.2 \times 10^{-16}$, Levene's test, non-overlapping 1 Mb windows) (Supplemental Fig. S10A,B). Similarly, the s.d. of $\log_2$ copy number changes was 0.11 ($\pm 8.2 \times 10^{-7}$) for the hamster genome at 75 nM compared to 0 nM paclitaxel (6 weeks growth) and 1.2 ($\pm 5.5 \times 10^{-6}$) for the human (F[1,4823] = 2758, $P < 2.2 \times 10^{-16}$, Levene's test, non-overlapping 1 Mb windows) (Supplemental Fig. S10C,D).

8 Statistical models for human and hamster genome scans

8.1 Negative binomial mixed model

Genes that confer a survival advantage as a result of growth or paclitaxel will show increased read counts and vice versa. We identified human loci for growth and paclitaxel by evaluating the significance of read changes across the genome using a negative binomial mixed model to control for both overdispersion (Section 8.4) and batch effects of RH pool (Supplemental Fig. S11)

$$log(Y^{\mathrm{Hu}}_{ijkl}) \sim log(N^{\mathrm{Hu}}_{ijkl}) + \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \quad (3)$$

where Y^{Hu}_{ijkl} is the number of species-specific human reads in each 1 Mb window and N^{Hu}_{ijkl} is the total number of species-specific aligned human reads in each sample, with its logarithm representing the offset to correct for differences in sequence ascertainment across samples. The other symbols are as described for Equation 2.

To identify additional human regulatory loci regulating the abundance of hamster mitochondria, human mitochondria or *M. fermentans*, relevant read numbers were added individually to the model as fixed effects (Equations 7, 9 and 12). The model described in Equation 2 was also modified to identify loci for growth and paclitaxel in the hamster genome (Equation 10). Five sets of genome scans were thus performed.

The mixed models were analyzed using the gam function of the mgcv package (Wood 2017) employing restricted maximum likelihood (REML) estimation. We obtained P values using the glht function of the multcomp package (Hothorn et al. 2008). Growth loci were identified employing the null hypotheses of no copy changes at 0, 8, 25 and 75 nM paclitaxel and an average conditional effect at 27 nM paclitaxel. Paclitaxel loci were identified using null hypotheses of no copy changes at 1, 2, 3, 4 and 6 weeks and an average conditional effect at 3.2 weeks. Interaction P values were also calculated. Each set of genome scans took ∼2 h on a computer cluster.

Genome scans, related statistical analyses and graphical representations used overlapping 1 Mb windows with 10 kb steps. When noted, other statistical tests employed non-overlapping 1 Mb windows to be conservative. All errors in text are standard errors of the mean (s.e.m.), unless otherwise remarked.

8.2 Significance thresholds

Genome-wide significance thresholds were set by permutation to give a 5 % family-wise error rate (FWER), equivalent to one expected false positive signal every 20 genome scans (Churchill, Doerge 1994).

Significance thresholds for human or hamster loci were obtained by co-shuffling $Y^{\mathrm{Hu\ or\ Ha}}_{ijkl}$ and $N^{\mathrm{Hu\ or\ Ha}}_{ijkl}$ in Equation 3 within growth (G_i), drug (D_j) and pools (P_k), thus keeping the relationship between $Y^{\mathrm{Hu\ or\ Ha}}_{ijkl}$ and $N^{\mathrm{Hu\ or\ Ha}}_{ijkl}$ intact within the fixed and random effects. More than 1000 shuffles of the genome were performed for all overlapping 1 Mb windows, taking ∼25 days on a computer cluster for each of the two sets of genome scans. The maximum $-\log_{10} P$ values were chosen from the shuffle of each genome scan and genome-wide significance thresholds obtained by choosing the 95$^{\mathrm{th}}$ percentile of the 1000 maximum values.

Significance thresholds for human loci regulating the abundance of hamster mitochondria, human mitochondria or *M. fermentans* were obtained by permuting Y^{Hu}_{ijkl} and N^{Hu}_{ijkl} within P_k. The 1000 shuffles took ∼15 days for each of the three sets of genome scans.

8.3 Peak finding

Human loci were identified using an automated peak finding algorithm that allowed the minimum distance between peaks to be adjusted (Supplemental Tables S8, S11). Comparison with handpicked growth loci on Chromosomes 1–3 (0 nM paclitaxel) showed an optimum concordance of 68 % ± 6 % when the minimum

distance between peaks was set at 2 Mb. The algorithm was more conservative (35 ± 0.6 loci, first three chromosomes) than manual peak finding (46 ± 3.5 loci), but this difference was not significant $t[1,2.1] = 3.2$, $P = 0.08$, Welch's Two Sample t-test).

8.4 Model fit for human loci

There was significant overdispersion of the data used to identify human growth and paclitaxel loci (dispersion = 31.7 ± 0.4 Poisson model, non-overlapping 1 Mb windows). Using the negative binomial mixed model moved the dispersion parameter significantly closer to unity (dispersion = $0.96 \pm 4.0 \times 10^{-3}$ negative binomial; $t[1,3113] = 69$, $P < 2.2 \times 10^{-16}$, non-overlapping 1 Mb windows).

The fit of the negative binomial mixed model was significantly improved by the random effects (AIC = 1311 ± 6 without random effects, 1182 ± 5 with random effects, $t[1,6182] = 17$, $P < 2.2 \times 10^{-16}$, non-overlapping 1 Mb windows) (Supplemental Fig. S11). Both the pool ($\chi^2 = 874 \pm 14$, reference df = 5, $-\log_{10} P = 18.0 \pm 0.3$, non-overlapping 1 Mb windows) and drug history random effects were significant ($\chi^2 = 226 \pm 4$, reference df = 28, $-\log_{10} P = 2.5 \pm 0.03$, non-overlapping 1 Mb windows).

Inspection of residuals in eight randomly selected genome positions suggested satisfactory performance of the model, including control of overdispersion and RH pool batch effects (Supplemental Figs. S12–S17). These conclusions were supported by a goodness-of-fit analysis for the whole dataset, which showed no significant departure of the residuals from normal ($\chi^2 = 92.9 \pm 0.4$, df = 93.6 ± 0.1, $-\log_{10} P = 0.43 \pm 2.1 \times 10^{-3}$, likelihood ratio test, non-overlapping 1 Mb windows).

9 Reproducibility

To assess replicability, we used a cross-validation procedure in which non-overlapping halves of the six RH pools were compared using the intersection of significant 1 Mb windows at various false discovery rate (FDR) thresholds (Supplemental Fig. S18) (Benjamini, Hochberg 1995). Highly significant overlaps were found, suggesting good reproducibility (summary logarithm of the odds ratio = 0.56 at FDR = 0.01, 95 % confidence intervals = 0.37 and 0.75, $t[1,10.0] = 5.7$, $P = 2.0 \times 10^{-4}$, Kenward-Roger df).

10 Further details on growth loci

10.1 Growth loci tally

The number of growth loci at the five paclitaxel levels ranged between 55 and 460, giving a total of 1836 loci (Fig. 4A; Supplemental Fig. S19). However, there was significant overlap between the growth loci at the various paclitaxel concentrations (Supplemental Fig. S20; Supplemental Tables S9, S10). There were 859 unique growth loci after selecting representatives with the highest $-\log_{10} P$ values, ∼1.4 % of all known genes (Supplemental Table S8). References in text and figures are to unique growth loci, unless otherwise noted.

Seven of the 859 unique growth loci (*SEMA3A*, *TOR1A*, *ATXN8OS*, *RN7SL584P*, *DNAH17*, *DNAH17-AS1* and *AL050309.1*) had positive growth coefficients (0.11 ± 0.02); the remaining coefficients were negative ($-0.24 \pm 2.4 \times 10^{-3}$, compared to -0.15 ± 0.0 in all non-overlapping 1 Mb windows) (Supplemental Fig. S23A). These observations are consistent with the general loss of human DNA in the RH samples as a result of growth (Fig. 3A–D; Supplemental Figs. S7–S10).

10.2 Overlap of growth loci

There were significant correlations between the $-\log_{10} P$ values for growth at different paclitaxel concentrations (weakest correlation, 0 nM and 75 nM; $R = 0.21$, $t[1,3052] = 12.134$, $P < 2.2 \times 10^{-16}$; mean

$R = 0.73 \pm 0.09$; non-overlapping 1 Mb windows) (Supplemental Fig. S20; Supplemental Table S9). However, the average correlation was decreased for the 75 nM samples ($R = 0.44 \pm 0.10$) compared to the others ($R = 0.93 \pm 0.02$; t[1,3.3] = 4.8, $P = 0.01$, Welch's Two Sample t-test).

The generally high similarity of $-\log_{10} P$ values for growth at different paclitaxel concentrations was reflected in the sharing of significant non-unique growth loci (least significant overlap, 0 nM and 75 nM; odds ratio = 8.5, $P = 6.2 \times 10^{-3}$, Fisher's Exact Test) (Supplemental Fig. S20; Supplemental Table S10). Overall, there was decreased overlap of loci at 75 nM (odds ratio = 33 ± 13) compared to the other levels (odds ratio = 1799 ± 1441), but the difference was insignificant (t[1,5.0] = 1.2, $P = 0.28$, Welch's Two Sample t-test).

10.3 Growth centromeres

Four of the unique growth loci (0.5 %) mapped to centromeres on Chromosomes 1, 11, 15 and X. These loci had significantly lower $-\log_{10} P$ values than the unique non-centromeric loci (centromere $-\log_{10} P = 19.3 \pm 1.3$, non-centromere $= 23.7 \pm 0.5$, t[1,4.0] = 3.2, $P = 0.03$, Welch's Two Sample t-test).

10.4 Coding and non-coding growth loci

Of the 855 unique non-centromeric RH growth loci, 442 (52 %) were coding and 413 (48 %) were non-coding, a significant enrichment in coding genes compared to the null of all genome positions (44 % coding, 56 % non-coding; $\chi^2[1] = 19.5, P = 9.9 \times 10^{-6}$). However, there was no significant difference in $-\log_{10} P$ values for the unique coding and non-coding loci (mean $-\log_{10} P = 23.5 \pm 0.6$ for coding, 24.0 ± 0.8 for non-coding; t[1, 825] = 0.4, $P = 0.7$, Welch's Two Sample t-test). These observations suggest that while coding genes are enriched, non-coding genes still make appreciable contributions to cell growth.

Close-ups of growth loci at the chromosome level are shown in Fig. 5A–C. Read count changes are shown for six significant loci in Fig. 5D. Loci displayed sharp mapping resolution, with the $-2\log_{10} P$ distance being $\lesssim$100 kb (Fig. 5E–F). However, loci that mapped to large genes ($\geqslant$ 1 Mb) tended to display a plateau at the top of their $-\log_{10} P$ peak, a testament to a high resolution mapping technique that can "trace" the profile of a lengthy gene (Supplemental Fig. S22).

10.5 Chromosomal distribution of growth loci

The number of unique growth loci (excluding centromeres) compared to the number of genes on each chromosome is shown in Supplemental Fig. S21A. The distribution of loci was non-uniform ($\chi^2[22] = 75, P = 9.8 \times 10^{-8}$), with significant surfeits on Chromosomes 2, 4–5 and X (least significant, Chromosome X; $\chi^2[1] = 4.0, P = 0.046$) and significant deficits on Chromosomes 14, 17 and 19–21 (least significant, Chromosome 20; $\chi^2[1] = 4.7, P = 0.03$).

10.6 Functional enrichment of RH growth genes

The unique RH growth genes were enriched in ten categories of the functional annotation chart of DAVID including alternative splicing, splice variant and phosphoprotein (FDR < 0.05) (Benjamini, Hochberg 1995; Huang et al. 2009). Further, 31 terms of the biological process category of the gene ontology (GO) were enriched, including neurogenesis, embryonic development and cellular biosynthesis (FDR < 0.05) (Mi et al. 2019; The Gene Ontology Consortium 2019). The unique RH growth genes were also significantly enriched in the category of cell adhesion in both DAVID (observed = 24, expected = 10, odds ratio = 2.5, $P = 2.5 \times 10^{-4}$, EASE modified Fisher's Exact Test, FDR = 0.02) and GO (observed = 43, expected = 19, odds ratio = 2.5, $P = 1.7 \times 10^{-6}$, Fisher's Exact Test, FDR = 4.4×10^{-3}), a possible consequence of selection for adherent cells. Essential coding genes identified in two loss-of-function CRISPR screens were

enriched in housekeeping categories related to cell division, transcription, splicing and translation (Hart et al. 2015; Wang et al. 2015), overlapping with the RH growth genes in the splicing terms.

10.7 Novel RH growth genes

The unique coding RH growth genes had significantly fewer literature citations in the GeneRIF database (mean citations = 398 ± 15) than coding non-growth genes (896 ± 72; $t[1{,}164] = 6.8$, $P = 2.3 \times 10^{-10}$, Welch's Two Sample t-test at peak significance threshold of 364 publications, FDR = 2.3×10^{-9}) (Supplemental Fig. S28A,B) (Jimeno-Yepes et al. 2013). Similarly, there were fewer entries in the Reactome database for unique coding RH growth genes (mean citations = 249 ± 20) than coding non-growth genes (400 ± 20), which reached nominal but not FDR corrected significance ($t[1{,}4.3] = 5.4$, $P = 4.7 \times 10^{-3}$, Welch's Two Sample t-test at peak significance threshold of 224 entries, FDR = 0.09) (Supplemental Fig. S28C,D) (Fabregat et al. 2018).

Of the 442 unique coding region RH growth genes, 148 had neither literature citations in GeneRIF nor entries in the Reactome database. These genes showed significant enrichment of one protein domain, the WW domain (observed = 6, expected = 0.4, odds ratio = 18.9, $P = 3.3 \times 10^{-5}$, EASE modified Fisher's Exact Test, FDR = 9.6×10^{-3}), as judged by the InterPro database (Huang et al. 2009; Mitchell et al. 2019).

None of the 413 unique non-centromeric non-coding RH genes had entries in either the GeneRIF or Reactome databases.

10.8 Olfactory receptor genes

Reports have suggested that olfactory receptors may play a role in cell proliferation and differentiation (Kang, Koo 2012; Tsai et al. 2017). The unique RH growth loci were found in olfactory receptor gene clusters on six chromosomes: 3, 11, 12, 14, 19 and X. One locus with $-\log_{10} P = 21.5$ mapped within the Chromosome 19 cluster to *OR7C1* (*AC005255.1*, GENCODE v31; Supplemental Fig. S29A), an olfactory receptor which has been shown to play a role in maintenance of colon cancer cells (Morita et al. 2016). Although 19 out of 298 olfactory receptor genes displayed a growth phenotype in at least one of the five cell lines in a recent CRISPR screen (Hart et al. 2015), none overlapped with the RH growth genes. This observation suggests complementary gain- and loss-of-function roles for olfactory receptors in growth.

10.9 Gene deserts

"Gene deserts" are defined as regions of the genome >500 kb that lack known genes (Salzburger et al. 2009). The majority of unique RH growth loci mapped close to a gene (mean distance = -1.7 kb $\pm$ 1.4 kb). However, two growth loci were located in gene deserts, with the nearest gene > 250 kb away (Supplemental Fig. S29B,C).

The RH gene desert growth locus at 104 730 000 bp on Chromosome 1 (Supplemental Fig. S29B) was 11 897 bp away from variant rs10494021 ($R^2 \sim 1$) (1000 Genomes Project Consortium et al. 2015). This polymorphism had suggestive significance (odds ratio = 1.14, P = 6.61×10^{-6}) in a GWAS for childhood ear infection (Tian et al. 2017). Thus, a possible candidate for this clinical trait may be a novel RH gene rather than a known, but distant, gene regulated by rs10494021.

No significant difference was found in the $-\log_{10} P$ values for the unique growth loci in gene deserts (mean = 45.5 ± 29.5) compared to the other non-centromeric growth loci (mean = 23.7 ± 0.5; $t[1{,}1.0] = 0.74$, $P = 0.59$, Welch's Two Sample t-test). Although test power was low because there were only two RH gene desert loci, the lack of difference in $-\log_{10} P$ values lends some credence to these unusual loci.

11 Further details on paclitaxel loci

11.1 Paclitaxel loci tally

There were a total of 97 paclitaxel loci (Fig. 4B; Supplemental Figs. S20, S30; Supplemental Tables S9–S11). After accounting for overlap at the various growth times and selecting representatives with the highest $-\log_{10} P$ values, there were 38 unique paclitaxel loci (Supplemental Table S11). References in text and figures are to unique paclitaxel loci, unless otherwise noted.

Seven of the 38 unique paclitaxel loci (*NEK10*, *RPL7A*, *DNAJC1*, *TBC1D12*, *ZBED5-AS1*, *GALNT18*, *AC013762.1*) had positive coefficients ($1.7 \times 10^{-2} \pm 2.4 \times 10^{-3}$) indicating increased read counts as a result of drug exposure. The remaining unique paclitaxel loci had negative coefficients ($-2.1 \times 10^{-2} \pm 1.7 \times 10^{-3}$; null of all non-overlapping 1 Mb windows, $-2.0 \times 10^{-4} \pm 4.5 \times 10^{-5}$) (Supplemental Fig. S23B). The proportion of unique paclitaxel loci with positive coefficients was significantly greater than the unique growth loci (seven out of 859 unique growth loci; log-likelihood ratio = −15.8, $P = 2.4 \times 10^{-8}$, Multinomial Goodness-Of-Fit test).

11.2 Overlap of paclitaxel loci

There were significant correlations between $-\log_{10} P$ values for paclitaxel at different times (Supplemental Fig. S20; Supplemental Table S9) (weakest significant correlation, week 1 and average conditional effect, $R = 0.075$, $t[1{,}3052] = 4.2$, $P = 3.2 \times 10^{-5}$; mean $R = 0.62 \pm 0.10$, including insignificant comparisons between week 1 and weeks 4 or 6; non-overlapping 1 Mb windows).

There was a significant increase in the number of non-unique paclitaxel loci with time, increasing from zero at week 1 to 25 at week 6 ($\chi^2[4] = 43.7$, P $= 7.4 \times 10^{-9}$) (Fig. 4B; Supplemental Fig. S30; Supplemental Table S11). As expected from the significant correlations of $-\log_{10} P$ values, there was also significant overlap of the paclitaxel loci at the different times (least significant overlap, weeks 2 and 3, odds ratio = ∞, $P = 1.8 \times 10^{-7}$, Fisher's Exact Test) (Supplemental Fig. S20; Supplemental Tables S9, S10). However, there were also non-overlapping loci, suggesting that the various drug exposure times bring different genes into play (Khan et al. 2016; Wang, Kruglyak 2014).

11.3 Shared RH growth and paclitaxel loci

We evaluated the correlations between the $-\log_{10} P$ values for growth and paclitaxel (Supplemental Fig. S33A; Supplemental Table S9). As expected, the correlations were significantly higher for growth at 75 nM paclitaxel ($R = 0.27 \pm 0.06$) than growth at other paclitaxel levels ($R = 0.01 \pm 0.02$, non-overlapping 1 Mb windows, $t[1{,}6.6] = 3.7$, $P = 8.8 \times 10^{-3}$, Welch's Two Sample t-test).

There was also significant sharing between the 38 unique paclitaxel loci and 859 unique growth loci with 10 loci in common (odds ratio = 25.1, $P = 1.0 \times 10^{-10}$, Fisher's Exact Test) (Supplemental Fig. S32). As expected, the overlap between unique growth and paclitaxel genes was greatest at the highest growth time (week 6) and paclitaxel concentration (75 nM) (odds ratio = 711, $P = 2.2 \times 10^{-16}$, Fisher's Exact Test).

Similar results were obtained for the non-unique growth and non-unique paclitaxel loci, with the highest overlap at week 6 growth and 75 nM paclitaxel (odds ratio = 490, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test) (Supplemental Fig. S33B; Supplemental Table S10).

11.4 Paclitaxel centromeres

Four of the 38 unique paclitaxel loci were centromeres, located on Chromosomes 11, 16, 20 and X. There was no significant difference in $-\log_{10} P$ values for the unique centromeric ($-\log_{10} P = 26.3 \pm 5.6$) and

non-centromeric paclitaxel loci ($-\log_{10} P = 16.6 \pm 1.2$; t[1,3.3] = 1.7, $P = 0.18$, Welch's Two Sample t-test).

The frequency of centromeres in the unique paclitaxel loci (four out of 38, 10.5 %) was significantly higher than for the unique growth loci (four centromeres out of 859 loci, 0.5 %; log-likelihood ratio = −8.8, $P = 3.1 \times 10^{-5}$, Multinomial Goodness-Of-Fit test).

There were two centromeres in common between the unique growth and unique paclitaxel loci, on Chromosomes 11 and X. This overlap was insignificant likely because of the low numbers (odds ratio = 7.4, $P = 0.12$, Fisher's Exact Test). There was no significant difference between the $-\log_{10} P$ values of the unique paclitaxel centromeres ($-\log_{10} P = 26.3 \pm 5.6$) and unique growth centromeres ($-\log_{10} P = 19.3 \pm 1.3$; t[1,3.3] = 1.2, $P = 0.31$, Welch's Two Sample t-test).

11.5 Coding and non-coding paclitaxel loci

The nearest genes were coding for 19 (56 %) and non-coding for 15 (44 %) of the 34 non-centromeric unique paclitaxel loci. In contrast to the RH growth loci, which were enriched in coding genes, there was no significant difference in the numbers of unique coding and non-coding paclitaxel genes compared to randomly chosen locations in the genome ($\chi^2[1] = 1.9, P = 0.17$). Although the RH coding and non-coding growth genes had similar $-\log_{10} P$ values, the unique coding region paclitaxel loci had significantly lower $-\log_{10} P$ values (14.3 ± 1.5) than the non-coding (19.5 ± 1.9; t[1,28.5] = 2.1, $P = 0.040$, Welch's Two Sample t-test). Together, our observations suggest that coding and non-coding genes make important contributions to the action of paclitaxel as well as growth.

11.6 Chromosomal distribution of paclitaxel loci

The number of unique paclitaxel loci compared to the number of genes on each chromosome is shown in Supplemental Fig. S21B. The distribution of loci was not significantly different from uniform ($\chi^2[22] = 22.8, P = 0.4$), although post-hoc testing revealed one significant result, a surfeit on Chromosome 7 ($\chi^2[1] = 6.7, P = 9.7 \times 10^{-3}$).

11.7 Novel paclitaxel loci

Of the 19 unique coding region RH paclitaxel genes, eight had neither GeneRIF nor Reactome entries. Of the eight, *AC020915.5* also lacked entries in the PubMed database. None of the 15 unique non-coding non-centromeric RH paclitaxel genes had entries in either GeneRIF or Reactome. Study of these novel genes may provide new insights into paclitaxel action.

12 Further details on interaction loci

12.1 Interaction loci tally

The majority of the 62 interaction loci (60 out of 62) had negative coefficients ($-5.2 \times 10^{-3} \pm 1.7 \times 10^{-4}$; coefficients in all non-overlapping 1 Mb windows, $-8.9 \times 10^{-4} \pm 3.5 \times 10^{-5}$), while the other two loci, *LINC01487* and *LINC02752* (*AC111188.1*, GENCODE v31), had positive coefficients ($5.4 \times 10^{-3} \pm 3.4 \times 10^{-4}$) (Supplemental Figs. S23C, S30, S34).

12.2 Overlap of interaction loci with growth and paclitaxel loci

The interaction $-\log_{10} P$ values showed significantly higher correlation with paclitaxel ($R = 0.60 \pm 0.14$) than growth ($R = 0.03 \pm 0.14$, non-overlapping 1 Mb windows, t[1,8.9] = 2.8, $P = 0.02$, Welch's Two Sample t-test) (Supplemental Fig. S33A; Supplemental Table S9). Consistent with this observation, the

correlation of the interaction and growth $-\log_{10} P$ values at 75 nM paclitaxel ($R = 0.60$) was significantly higher than at the other paclitaxel concentrations ($R = -0.11 \pm 0.03$, non-overlapping 1 Mb windows, $t[1{,}3] = 9.1$, $P = 2.8 \times 10^{-3}$, Two Sample t-test with pooled variance).

There were 15 genes that overlapped between the interaction loci and the 859 unique growth loci (odds ratio $= 22.6, P = 8.3 \times 10^{-15}$, Fisher's Exact Test) and 14 genes that overlapped with the 38 unique paclitaxel loci (odds ratio $= 731, P < 2.2 \times 10^{-16}$, Fisher's Exact Test) (Supplemental Fig. S32). As expected, the overlap between the interaction loci and the unique growth loci was greatest at the highest paclitaxel concentration, 75 nM, (odds ratio = 586, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test), while the overlap between the interaction loci and the unique paclitaxel loci was greatest at the highest growth time, 6 weeks (odds ratio = 1522, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test).

A total of 22 interaction loci were shared with either the unique growth or unique pacliatexel loci (odds ratio = 38, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test), while a total of 7 interaction loci were shared with both (odds ratio = 2447, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test) (Supplemental Fig. S32).

Similar overlap results were obtained for the interaction and non-unique growth and paclitaxel loci. The highest overlap between the interaction and non-unique growth loci was at 75 nM paclitaxel (odds ratio = 482, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test) and the highest overlap between the interaction and non-unique paclitaxel loci was at 6 weeks of growth (odds ratio = 1522, $P < 2.2 \times 10^{-16}$, Fisher's Exact Test) (Supplemental Fig. S33B; Supplemental Table S10).

12.3 Interaction centromeres

Five interaction loci mapped to centromeres on Chromosomes 11, 15, 16, 20 and X. There was no significant difference in $-\log_{10} P$ values between the centromeric ($-\log_{10} P = 16.2 \pm 2.2$) and non-centromeric interaction loci ($-\log_{10} P = 12.0 \pm 0.5$; $t[1{,}4.3] = 1.9$, $P = 0.13$, Welch's Two Sample t-test).

Three of the five interaction centromeres overlapped with the four unique growth centromeres (Chromosomes 11, 15 and X; odds ratio = 20, $P = 0.021$, Fisher's Exact Test) and four with the four unique paclitaxel centromeres (Chromosomes 11, 16, 20 and X; odds ratio $= \infty$, $P = 5.6 \times 10^{-4}$, Fisher's Exact Test). There was no significant difference between the $-\log_{10} P$ values of the interaction centromeres (16.2 ± 2.2) and either the unique growth centromeres (19.3 ± 1.3, $t[1{,}6.2] = 1.2$, $P = 0.26$, Welch's Two Sample t-test) or the unique paclitaxel centromeres (26.3 ± 5.6, $t[1{,}3.9] = 1.7$, $P = 0.17$, Welch's Two Sample t-test).

12.4 Coding and non-coding interaction loci

The nearest genes were coding for 26 (46 %) and non-coding for 31 (54 %) of the 57 non-centromeric interaction loci. There was no significant difference in the proportion of coding and non-coding interaction genes compared to randomly chosen locations in the genome ($\chi^2[1] = 0.05, P = 0.83$). In addition, there were no significant differences in the $-\log_{10} P$ values between the coding ($-\log_{10} P = 12.4 \pm 0.9$) and non-coding interaction loci ($-\log_{10} P = 11.6 \pm 0.4$; $t[1{,}36.5] = 0.8$, $P = 0.4$, Welch's Two Sample t-test). Together, our observations suggest that coding and non-coding genes make important contributions to the interaction between growth and paclitaxel.

12.5 Chromosomal distribution of interaction loci

The number of interaction loci compared to the number of genes on each chromosome is shown in Supplemental Fig. S21C. The distribution of loci was significantly different from uniform ($\chi^2[22] = 50.9, P = 4.4 \times 10^{-4}$), with significant surfeits on Chromosomes 6 ($\chi^2[1] = 18.2, P = 1.9 \times 10^{-5}$) and 13 ($\chi^2[1] = 5.5, P = 0.02$).

12.6 Novel interaction loci

Of the 25 coding region RH interaction genes, five had neither GeneRIF nor Reactome entries. However, all of these five genes had entries in the PubMed database. None of the 31 non-coding non-centromeric RH interaction genes had entries in either GeneRIF or Reactome.

13 Public data

The human gene list was from the comprehensive annotation set of GENCODE v31 (GRCh38.p12) (Frankish et al. 2019). Loss-of-function growth genes identified in CRISPR screens (Hart et al. 2015; Wang et al. 2015) and microarray expression data from the G3 human RH panel (Wang et al. 2011) were from published studies. Human tissue RNA-Seq data were retrieved from GTEx v8 (The GTEx Consortium 2015).

Paralogs were identified using the duplicated genes database (DGD), employing Ensembl release 71 (Ouedraogo et al. 2012). The degree of intolerance to predicted loss-of-function (pLoF) variation in human genes was measured using the observed/expected ratio from the Genome Aggregation Database (gnomAD) release 2.1.1. (Karczewski et al. 2020). The ratio of nonsynonymous to synonymous substitutions (dN/dS) was calculated using mouse-human homologs in Ensembl release 97 (Cunningham et al. 2019).

Homologene release 68 was employed to evaluate the number of species with gene orthologs (Sayers et al. 2019). Protein-protein interactions were identified using STRING v11 (Szklarczyk et al. 2019).

Functional analysis of RH genes was performed using DAVID v6.8 (Huang et al. 2009). Enrichment in multisubunit protein complexes was assessed using the GO_CC_ALL category of DAVID, as described (Hart et al. 2015). GO analyses used the 2019-07-03 release and the PANTHER Overrepresentation Test (Mi et al. 2019; The Gene Ontology Consortium 2019).

InterPro was employed to assess protein domains (Huang et al. 2009; Mitchell et al. 2019). Coding genes of unknown function were classified as those lacking entries in GeneRIF (Jimeno-Yepes et al. 2013) and the Reactome Knowledgebase v70 (Fabregat et al. 2018).

14 Further details on mapping resolution

14.1 Mapping resolution using *TK1*

A linear mixed model was used to evaluate the distance between the location of the *TK1* gene and its retention peak in the RH samples. The model employed fixed and interaction effects of growth time and paclitaxel concentration, combined with random intercepts of RH pool and nested drug history

$$\Delta_{ijkl}^{TK1} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \tag{4}$$

where Δ_{ijkl}^{TK1} is the distance between the location of the *TK1* gene and its retention peak (bp). The other symbols, R packages and post hoc analyses are as described for Equation 2.

There were no significant fixed effects of growth ($t[1{,}92.5] = 0.2, P = 0.9$, Kenward-Roger df; average conditional effect) or drug concentration ($t[1{,}25.6] = 1.8, P = 0.08$, Kenward-Roger df; average conditional effect), but the interaction was significant ($t[1{,}94.7] = 2.7, P = 9.4 \times 10^{-3}$, Kenward-Roger df). These observations indicate that, as expected, growth time or drug concentration had minimal effects on *TK1* localization. There was a significant random effect of pool ($\chi^2[1] = 7.7, P = 5.4 \times 10^{-3}$, ICC = 0.31, likelihood ratio test) but not drug history ($\chi^2[1] = 3.0, P = 0.09$, ICC = 0.12, likelihood ratio test), consistent with the observed batch effect of RH pool (Supplemental Fig. S11).

The distance between the peak retention of *TK1* and the location of the gene was estimated as -27 kb $\pm$ 66 kb (average conditional effects) using the mixed model. This distance was statistically

indistinguishable from zero (t[1,5.1] $= 0.4, P = 0.7$, Kenward-Roger df; average conditional effects) (Supplemental Fig. S35A,C).

14.2 Mapping resolution using centromeres

As an independent approach to evaluating mapping accuracy, we measured the distances between the observed retention peak for each centromere and their assigned positions in the genome (Fig. 2D–F; Supplemental Figs. S5, S6, S35B,D,E). The functional locations of centromeres are poorly defined, and generous limits are provided on their positions in the human genome build GRCh38/hg38 (https://genome.ucsc.edu). For this analysis, we took the consensus locations of the centromeres to be the center of the contigs covering each centromere in hg38.

Some centromeres showed clearly defined retention peaks, while others showed a plateau, suggesting uniformly dispersed functional attributes (Fig. 2E,F; Supplemental Fig. S6). Nevertheless, we reasoned that certain centromere sequences will have greater functional importance than others, as indicated by the maximum retention peak. To evaluate the distances between the observed retention peak for each centromere and their consensus assignations in the genome, we used the linear mixed model described for *TK1* (Equation 4), with the addition of a random intercept of chromosome

$$\Delta^{\mathrm{CEN}}_{ijklm} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + Chr_m + \epsilon_{ijklm}, \tag{5}$$

where $\Delta^{\mathrm{CEN}}_{ijklm}$ is the distance between the centromere retention peaks and their assigned locations in the genome (bp) and *Chr* is the random intercept due to chromosome m = 1–22, X. The other symbols, R packages and post hoc analyses are as described for Equation 2.

Similar to *TK1*, the fixed effects of growth (t[1,2293] $= 0.7, P = 0.5$, Kenward-Roger df; average conditional effect), paclitaxel (t[1,27.6] $= 0.06, P = 0.95$, Kenward-Roger df; average conditional effect) and their interaction (t[1,2053] $= 0.05, P = 0.96$, Kenward-Roger df) had no significant effect on centromere mapping accuracy. Also as for *TK1*, and consistent with the observed RH pool batch effects (Supplemental Fig. S11), the random effect of pool was significant ($\chi^2[1] = 4.5, P = 0.03$, likelihood ratio test), but the nested random effect of drug history was not ($\chi^2[1] = 0.09, P = 0.76$, likelihood ratio test). Nevertheless, the variance due to the RH pools and drug history was minor, accounting for only 0.50 % and 0.10 % of the variance components, respectively (ICC $= 5.0 \times 10^{-3}$, 1.0×10^{-3}).

In contrast, the between-chromosome variance dominated the RH pool effect, accounting for 37 % of the variance components (ICC = 0.37). Further, the random effect of chromosome ($\chi^2[1] = 1077, P < 2.2 \times 10^{-16}$, likelihood ratio test) was much more significant than that of the RH pools (Supplemental Fig. S35B). The significant chromosome random effect suggests that there are substantial differences between the functional locations of individual centromeres and their currently assigned positions. Our approach may allow more accurate localization of active centromere sequences.

The mean distance between the centromere retention peaks and their assigned genomic location estimated from the mixed model was statistically indistinguishable from zero (-30 kb $\pm$ 105 kb; t[1,24] $= 0.3, P = 0.8$, Kenward-Roger df; average conditional effects) (Supplemental Fig. S35D,E). The mapping accuracy using centromeres was similar to that obtained using *TK1*, together suggesting a resolution $\sim$30 kb. Despite the good mapping resolution of RH-BSA, regions of high gene density were not always clearly resolved (Supplemental Fig. S36). Higher radiation doses and pool numbers can help dissect such regions in the future.

15 Further details on mitochondrial copy number

15.1 Calculating mitochondrial copy number of HEK293 and A23 cells

Mitochondrial copy numbers were calculated by comparing the number of mitochondrial sequence reads with those from the nuclear genome, each normalized to their respective genome lengths. The mitochondrial copy number in HEK293 cells was thus calculated as $C^{\mathrm{M}}_{\mathrm{Hu}} = 3(N^{\mathrm{M}}_{\mathrm{Hu}}/N^{\mathrm{G}}_{\mathrm{Hu}})(L^{\mathrm{G}}_{\mathrm{Hu}}/L^{\mathrm{M}}_{\mathrm{Hu}})$, where $N^{\mathrm{M}}_{\mathrm{Hu}}$ is the number of human mitochondrial reads, $N^{\mathrm{G}}_{\mathrm{Hu}}$ is the number of human genome reads, $L^{\mathrm{G}}_{\mathrm{Hu}}$ is the length of the human genome (3 088 269 832 bp) and $L^{\mathrm{M}}_{\mathrm{Hu}}$ is the length of the human mitochondrial genome (16 569 bp). The factor of three was used to correct for the fact that HEK293 cells are thought to be pseudotriploid (Lin et al. 2014), while mitochondria are haploid (Supplemental Table S4).

The mitochondrial copy number in A23 cells was calculated as $C^{\mathrm{M}}_{\mathrm{Ha}} = 2(N^{\mathrm{M}}_{\mathrm{Ha}}/N^{\mathrm{G}}_{\mathrm{Ha}})(L^{\mathrm{G}}_{\mathrm{Ha}}/L^{\mathrm{M}}_{\mathrm{Ha}})$, where $N^{\mathrm{M}}_{\mathrm{Ha}}$ is the number of hamster mitochondrial reads, $N^{\mathrm{G}}_{\mathrm{Ha}}$ is the number of hamster genome reads, $L^{\mathrm{G}}_{\mathrm{Ha}}$ is the length of the hamster genome (2 368 906 908 bp) and $L^{\mathrm{M}}_{\mathrm{Ha}}$ is the length of the hamster mitochondrial genome (16 283 bp). The factor of two was was used to correct for the haploid and diploid nature of the mitochondrial and hamster genomes, respectively.

15.2 Calculating mitochondrial copy number of RH samples

The hamster mitochondrial copy number in the RH samples, including the RH pools, was calculated as $C^{\mathrm{M}}_{\mathrm{Ha}} = 2(N^{\mathrm{M}}_{\mathrm{Ha}}/N^{\mathrm{G}}_{\mathrm{Ha}})(L^{\mathrm{G}}_{\mathrm{Ha}}/L^{\mathrm{M}}_{\mathrm{Ha}})$ (Supplemental Table S12). The human mitochondrial copy number was calculated as $C^{\mathrm{M_G}}_{\mathrm{Hu}} = 2(N^{\mathrm{M}}_{\mathrm{Hu}}/N^{\mathrm{G}}_{\mathrm{Ha}})(L^{\mathrm{G}}_{\mathrm{Ha}}/L^{\mathrm{M}}_{\mathrm{Hu}})$. Symbols for both equations are defined in Section 15.1.

Human mitochondrial copy numbers were also estimated by comparing mitochondrial reads with human *TK1* reads and applying a correction factor for reversion (Supplemental Table S12). Thus, $C^{\mathrm{M}_{TK1}}_{\mathrm{Hu}} = (N^{\mathrm{M}}_{\mathrm{Hu}}/N^{TK1}_{\mathrm{Hu}})(L^{TK1}_{\mathrm{Hu}}/L^{\mathrm{M}}_{\mathrm{Hu}})(1-r)$, where N^{TK1}_{Hu} is the maximum number of human-specific reads in the 1 Mb window encompassing *TK1*, L^{TK1}_{Hu} is the length of the *TK1* window (1×10^6 bp) and r is the reversion frequency (Supplemental Table S1). The human *TK1* gene has a retention frequency of 1 in the RH samples and, like the mitochondrial genome, is essentially haploid. Therefore, no additional factor is required to correct for differences in ploidy.

15.3 Hamster mitochondrial copy number changes in the RH cells

We evaluated changes of the hamster mitochondrial copy number in the RH samples using a linear mixed model

$$C^{\mathrm{M}}_{\mathrm{Ha}_{ijkl}} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \tag{6}$$

where $C^{\mathrm{M}}_{\mathrm{Ha}_{ijkl}}$ is the hamster mitochondrial copy number (Section 15.2). The other symbols, R packages and post hoc analyses are as described for Equation 2.

Hamster mitochondrial copy numbers showed a significant decrease as a result of growth ($t[1{,}93.5] = 2.3, P = 0.02$, Kenward-Roger df; average conditional effect) (Supplemental Fig. S37A) and a significant increase as a result of paclitaxel ($t[1{,}28.1] = 3.4, P = 2 \times 10^{-3}$, Kenward-Roger df; average conditional effect) (Supplemental Fig. S37B). These changes are consistent with previous observations, which showed a decrease in mitochondria as a result of malignant growth (Reznik et al. 2016) and an increase in cells treated with paclitaxel (Karbowski et al. 2001). The interaction was insignificant ($t[1{,}95.5] = 1.5, P = 0.1$). There was a significant random effect of pool ($\chi^2[1] = 6.6, P = 0.01$, ICC = 0.18, likelihood ratio test), but not drug history ($\chi^2[1] = 0, P = 1$, ICC = 0, likelihood ratio test).

15.4 Details on human loci regulating hamster mitochondrial abundance

To identify human loci that regulated the levels of hamster mitochondria, we amended our negative binomial mixed model (Equation 3) by adding hamster mitochondrial reads as a Supplemental fixed effect

$$log(Y^{\mathrm{Hu}}_{ijkl}) \sim log(N^{\mathrm{Hu}}_{ijkl}) + \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + \beta_4 M^{\mathrm{Ha}}_{ijkl} + P_k + L_l + \epsilon_{ijkl}, \quad (7)$$

where Y^{Hu}_{ijkl} is the number of human-specific reads in each 1 Mb window and β_4 is the fixed effect of hamster-specific mitochondrial read number, M^{Ha}, at growth time i, paclitaxel concentration j, RH pool k and drug history l (Supplemental Fig. S38A–D). The other symbols, R packages and post hoc analyses are as described for Equation 3.

The addition of hamster mitochondrial abundance to the model resulted in only minor changes to the $-\log_{10} P$ values for growth, paclitaxel and their interaction (weakest correlation, paclitaxel week 1, $R = 0.84$, $t[1{,}3052] = 87$, $P < 2.2 \times 10^{-16}$; mean $R = 0.96 \pm 0.02$; non-overlapping 1 Mb windows).

15.5 Human mitochondrial copy number changes in the RH cells

Human mitochondrial copy numbers in the RH samples were calculated either by reference to the hamster nuclear genome ($C^{\mathrm{M_G}}_{\mathrm{Hu}}$) or the human *TK1* gene ($C^{\mathrm{M}_{TK1}}_{\mathrm{Hu}}$) (Section 15.2). The two methods gave comparable results.

Using reads from the hamster nuclear genome as a reference gave a copy number of 1.4 ± 0.8 for human mitochondria in the six RH pools, not significantly different from zero ($t[1{,}5] = 1.7$, $P = 0.14$; One Sample *t*-test) (Supplemental Fig. S37C,D; Supplemental Table S12). Using *TK1* as a reference gave a copy number of 2.8 ± 1.6 in the RH pools, also not significantly different from zero ($t[1{,}5] = 1.8$, $P = 0.14$; One Sample *t*-test) (Supplemental Fig. S37E,F; Supplemental Table S12).

In addition, the mean of the two methods gave a copy number of 2.1 ± 1.2 in the RH pools, which was not significantly different from zero ($t[1{,}5] = 1.8$, $P = 0.14$; One Sample *t*-test) (Supplemental Fig. S37G,H; Supplemental Table S12).

We modified the previously described linear mixed model (Equation 6) to evaluate the human mitochondrial copy number in all the RH samples

$$C^{\mathrm{M_G\ or\ M}_{TK1}}_{\mathrm{Hu}_{ijkl}} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \quad (8)$$

where $C^{\mathrm{M_G\ or\ M}_{TK1}}_{\mathrm{Hu}_{ijkl}}$ is the human mitochondrial copy number calculated by reference to the hamster nuclear genome or the human *TK1* gene, respectively. Other symbols and post hoc analyses are as described for Equation 2.

The results from the linear mixed model were of varying and marginal significance, depending on the measurement method used. The copy number in all samples was 0.6 ± 0.3 using the hamster nuclear genome as a reference ($t[1{,}5.2] = 2.1, P = 0.09$, Kenward-Roger df; average conditional effects), 1.4 ± 0.5 using *TK1* as a reference ($t[1{,}5.4] = 2.7, P = 0.04$, Kenward-Roger df; average conditional effects), and 1.0 ± 0.4 using the mean of the two methods ($t[1{,}5.3] = 2.7, P = 0.04$, Kenward-Roger df; average conditional effects).

There was no significant effect on human mitochondrial copy number of growth ($t[1{,}93.5] = 0.1, P = 0.9$, hamster alignments; $t[1{,}93.4] = 0.8, P = 0.4$, *TK1*; $t[1{,}93.5] = 0.7, P = 0.5$, mean of two methods; Kenward-Roger df, average conditional effects), paclitaxel ($t[1{,}28.1] = 0.4, P = 0.7$, hamster alignments; $t[1{,}27.2] = 1.8, P = 0.09$, *TK1*; $t[1{,}28.2] = 1.5, P = 0.2$, mean of two methods; Kenward-Roger df, average conditional effects) or their interaction ($t[1{,}95.5] = 1.6, P = 0.1$, hamster alignments; $t[1{,}95.5] = 0.9, P = 0.4$, *TK1*; $t[1{,}95.5] = 1.2, P = 0.2$, mean of two methods; Kenward-Roger df).

The random effects of pool were inconclusive ($\chi^2[1] = 10.8, P = 9.9 \times 10^{-4}$, ICC = 0.2, hamster alignments; $\chi^2[1] = 1.1, P = 0.3$, ICC = 0.07, *TK1*; $\chi^2[1] = 3.3, P = 0.07$, ICC = 0.1, mean of two methods; likelihood ratio tests), and drug history was insignificant ($\chi^2[1] = 0, P = 1$, ICC = 0, hamster

alignments; $\chi^2[1] = 0.2, P = 0.6$, ICC = 0.05, *TK1*; $\chi^2[1] = 0, P = 1$, ICC = 4.9×10^{-9}, mean of two methods; likelihood ratio tests).

15.6 Details on human loci regulating human mitochondrial abundance

Despite the low levels of human mitochondria, we nevertheless sought to identify human loci that regulated the abundance of these organelles by including the mitochondrial sequence reads as an additional fixed effect in our negative binomial mixed model (Equations 3 and 7)

$$log(Y^{\mathrm{Hu}}_{ijkl}) \sim log(N^{\mathrm{Hu}}_{ijkl}) + \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + \beta_4 M^{\mathrm{Hu}}_{ijkl} + P_k + L_l + \epsilon_{ijkl}, \quad (9)$$

where β_4 is the fixed effect of human-specific mitochondrial read number, M^{Hu}, at growth time i, paclitaxel concentration j, RH pool k and drug history l. The other symbols, R packages and post hoc analyses are as described for Equation 3.

The additional fixed effect left the $-\log_{10} P$ values for growth, paclitaxel and their interaction essentially unchanged from those obtained previously (weakest correlation, paclitaxel week 1, $R = 0.99$, t[1,3052] = 538, $P < 2.2 \times 10^{-16}$; mean $R = 0.998 \pm 4.9 \times 10^{-4}$; non-overlapping 1 Mb windows). Although no permutation significant loci were uncovered, we identified one human locus that regulated the levels of human mitochondria with FDR < 0.05 (Supplemental Fig. S38E,F).

16 Hamster genome changes

We evaluated whether hamster copy number alterations (CNAs) (Fig. 2A) confer selective advantages for growth and paclitaxel using the same statistical model we employed for the human genome (Equation 3)

$$log(Y^{\mathrm{Ha}}_{ijkl}) \sim log(N^{\mathrm{Ha}}_{ijkl}) + \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \quad (10)$$

where Y^{Ha}_{ijkl} is the number of hamster-specific reads in each 1 Mb window and N^{Ha}_{ijkl} is the total number of hamster-specific reads in each sample, with its logarithm representing the offset to correct for differences in sequence ascertainment across samples. The other symbols are as described for Equation 3.

Loci that exceeded genome-wide permutation significance thresholds were identified for growth, paclitaxel and their interaction (Supplemental Figs. S39, S40).

Despite the high quality of the hamster genome, a total of 341 contigs, many relatively short, still remained after discarding unanchored contigs (Rupp et al. 2018). Analyses of sequence alignment numbers in the hamster genome used all available contigs. However, for greater clarity and continuity in the hamster genome scans, we removed the 209 shortest remaining contigs (61 %; maximum length 3.6 Mb), leaving 132 contigs that encompassed 95 % of the hamster genome.

The greater abundance of hamster-specific reads compared to human (59 ± 14-fold, RH pools; 84 ± 16-fold, all samples; c.f. Supplemental Tables S5, S6 for RH pools), provided higher statistical power to detect loci. However, the $-\log_{10} P$ values for significant hamster growth loci (9.2 ± 0.1) were significantly less than for human (21.9 ± 0.2; t[1,3237] = 68, $P < 2.2 \times 10^{-16}$, Welch's Two Sample t-test; non-overlapping 1 Mb windows). A similar picture was evident for paclitaxel loci ($-\log_{10} P = 11.3 \pm 0.2$, hamster; 20.7 ± 1.0, human; t[1, 46] = 9.6, $P = 1.7 \times 10^{-12}$, Welch's Two Sample t-test; non-overlapping 1 Mb windows). The weaker $-\log_{10} P$ values of hamster loci are consistent with the greater stability of the hamster genome compared to human (Fig. 3A–D; Supplemental Fig. S10).

The mapping resolution using the hamster genome depends on the length of the CNAs, typically encompassing many megabases and genes. Further, the hamster genome is of high quality but is still divided into 132 contigs, even after pruning. Hamster CNAs are thus artefactually divided into smaller intervals. Nevertheless, examples of hamster growth and paclitaxel loci due to pre-existing CNAs are shown

in Supplemental Fig. S41. In contrast, CNAs in the donor HEK293 genome affect retention frequency but do not affect mapping resolution, which depends on the number of breakpoints caused by the radiation.

17 Further details on mycoplasma contamination

17.1 Quantitation of mycoplasma copy number

Mycoplasma copy numbers were calculated by comparing the number of mycoplasma reads with those from the host nuclear genome in each sample. For HEK293 cells the reference genome was human; for A23 cells and the RH samples the reference genome was hamster. Normalization was used to adjust for the respective lengths of each nuclear and mycoplasma genome. A further correction factor of two accounted for the diploid and haploid nature of the nuclear and mycoplasma genomes, respectively.

Thus, the mycoplasma copy number, $Mc = 2(N_{\mathrm{Mc}}/N^{\mathrm{Nu}}_{\mathrm{Host}})(L^{\mathrm{G}}_{\mathrm{Host}}/L^{\mathrm{G}}_{\mathrm{Mc}})$, where N_{Mc} is the number of sequence reads aligned to the mycoplasma genome, $N^{\mathrm{Host}}_{\mathrm{Nu}}$ is the number of reads aligned to the host nuclear genome, $L^{\mathrm{G}}_{\mathrm{Host}}$ is the length of the host nuclear genome and $L^{\mathrm{G}}_{\mathrm{Mc}}$ is the length of the relevant mycoplasma genome.

17.2 Mycoplasma contamination in parental cells

We examined the sequence data from the parental HEK293 and A23 cell lines to see if the cells were contaminated with mycoplasma. We aligned reads to the five species shown in Supplemental Table S13, with 1 bp and 4 bp mismatches giving identical read numbers.

A recent paper using publicly available RNA-Seq datasets evaluated the prevalence of mycoplasma contamination in published studies (Olarerin-George, Hogenesch 2015). Of the datasets, 11 % featured one or more contaminated samples, defined as ≥ 100 reads million^{-1} mapping to mycoplasma. None of our parental samples exceeded these levels (Supplemental Table S13). However, a lower threshold is appropriate for our dataset, since mammalian genome sequences are more complex than RNA-Seq datasets.

The HEK293 cells were nearly devoid of mycoplasma sequences. The A23 cells, however, had appreciable levels. The most abundant species was *M. fermentans* at 0.26 copies per cell, followed by *M. hominis* and *M. hyorhinis*.

17.3 Mycoplasma copy number in RH samples

To assess the effects of mycoplasma contamination in the RH samples, reads were aligned to *M. fermentans* using 1 bp mismatches. The *M. fermentans* copy number was 0.11 ± 0.03 in the six RH pools (t[1,5] = 3.4, $P = 0.02$; One Sample t-test).

We evaluated changes in the mycoplasma copy number in all RH samples by modifying the linear mixed model previously described in Equation 2

$$Mc_{ijkl} \sim \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + P_k + L_l + \epsilon_{ijkl}, \tag{11}$$

where Mc_{ijkl} is the mycoplasma copy number (Section 17.1). Other symbols and post hoc analyses are as described for Equation 2. The *M. fermentans* copy number in all samples was 0.13 ± 0.03 (t[1,5.16] = 4.9, $P = 4.1 \times 10^{-3}$, Kenward-Roger df; average conditional effects).

The mixed model revealed no significant effect of growth (t[1,91.5] = 1.1, $P = 0.3$, Kenward-Roger df; average conditional effect), paclitaxel (t[1,24.7] = 0.8, $P = 0.4$, Kenward-Roger df; average conditional effect) or their interaction (t[1,93.6] = 1.0, $P = 0.3$, Kenward-Roger df). However, there was a significant random effect of pool (χ^2[1] = 3.9, $P = 0.047$, ICC = 0.22, likelihood ratio test), and drug history (χ^2[1] = 13.0, $P = 3.2 \times 10^{-4}$, ICC = 0.25, likelihood ratio test).

17.4 Effects of mycoplasma copy number on genetic mapping

Although mycoplasma contamination is irrelevant to the original role of RH panels in high resolution genetic mapping, these bacteria could nevertheless affect the findings of our study. To evaluate the consequences of mycoplasma contamination on human loci, we modified our negative binomial mixed model (Equation 3) to include the *M. fermentans* reads as an additional fixed effect

$$log(Y^{\mathrm{Hu}}_{ijkl}) \sim log(N^{\mathrm{Hu}}_{ijkl}) + \beta_0 + \beta_1 G_i + \beta_2 D_j + \beta_3 G_i D_j + \beta_4 Mr_{ijkl} + P_k + L_l + \epsilon_{ijkl}, \qquad (12)$$

where β_4 is the fixed effect of the mycoplasma reads, Mr, at growth time i, paclitaxel concentration j, RH pool k and drug history l. The other symbols, R packages and post hoc analyses are as described for Equation 3.

The $-\log_{10} P$ values for growth and paclitaxel remained essentially unchanged (weakest correlation, paclitaxel week 1, $R = 0.97$, $t[1, 3052] = 222$, $P < 2.2 \times 10^{-16}$; mean $R = 0.995 \pm 2.4 \times 10^{-3}$; non-overlapping 1 Mb windows). In addition, there were no human loci that regulated the levels of the bacterium and that also exceeded either the permutation significance threshold or the more liberal FDR threshold of 0.05 (in fact, all FDR = 1). These observations indicate that the mycoplasma contamination had minimal effects on our conclusions.

18 Supplemental References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* **526:** 68–74. doi:10.1038/nature15393

Avner P, Bruls T, Poras I, Eley L, Gas S, Ruiz P, Wiles MV, Sousa-Nunes R, Kettleborough R, Rana A, et al. 2001. A radiation hybrid transcript map of the mouse genome. *Nat Genet* **29:** 194–200. doi:10.1038/ng1001-194

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Software* **67:** 1–48. doi:10.18637/jss.v067.i01

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57:** 289–300.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* **9:** 378–400. doi:10.32614/rj-2017-066

Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138:** 963–971.

Contreras-Galindo R, Fischer S, Saha AK, Lundy JD, Cervantes PW, Mourad M, Wang C, Qian B, Dai M, Meng F, et al. 2017. Rapid molecular assays to study human centromere genomics. *Genome Res* **27:** 2040–2049. doi:10.1101/gr.219709.116

Cox DR, Burmeister M, Price ER, Kim S, Myers RM. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250:** 245–250.

Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu S, et al. 2019. Ensembl 2019. *Nucleic Acids Res* **47:** D745–D751. doi:10.1093/nar/gky1113

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, et al. 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46:** D649–D655. doi:10.1093/nar/gkx1132

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47:** D766–D773. doi:10.1093/nar/gky955

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. 2014. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159:** 647–661. doi:10.1016/j.cell.2014.09.029

Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, et al. 2015. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163:** 1515–1526. doi:10.1016/j.cell.2015.11.015

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. *Biometrical J* **50:** 346–363.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57. doi:10.1038/nprot.2008.211

Hudson TJ, Church DM, Greenaway S, Nguyen H, Cook A, Steen RG, Van Etten WJ, Castle AB, Strivens MA, Trickett P, et al. 2001. A radiation hybrid map of mouse genes. *Nat Genet* **29:** 201–205. doi:10.1038/ng1001-201

Jimeno-Yepes AJ, Sticco JC, Mork JG, Aronson AR. 2013. GeneRIF indexing: sentence selection based on machine learning. *BMC Bioinformatics* **14:** 171. doi:10.1186/1471-2105-14-171

Kang N, Koo J. 2012. Olfactory receptors in non-chemosensory tissues. *BMB Rep* **45:** 612–622.

Karbowski M, Spodnik JH, Teranishi M, Wozniak M, Nishizawa Y, Usukura J, Wakabayashi T. 2001. Opposite effects of microtubule-stabilizing and microtubule-destabilizing drugs on biogenesis of mitochondria in mammalian cells. *J Cell Sci* **114:** 281–291.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581:** 434–443. doi:10.1038/s41586-020-2308-7

Khan AH, Bloom JS, Faridmoayer E, Smith DJ. 2016. Genetic screening reveals a link between Wnt signaling and antitubulin drugs. *Pharmacogenomics J* **16:** 164–172. doi:10.1038/tpj.2015.50

Kwitek AE, Gullings-Handley J, Yu J, Carlos DC, Orlebeke K, Nie J, Eckert J, Lemke A, Andrae JW, Bromberg S, et al. 2004. High-density rat radiation hybrid maps containing over 24,000 SSLPs, genes, and ESTs provide a direct link to the rat genome sequence. *Genome Res* **14:** 750–757. doi:10.1101/gr.1968704

Lenth R. 2019. emmeans: estimated marginal means, aka least-squares means. R package version 1.4.

Lin YC, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, Reumers J, Moisse M, Plaisance S, Drmanac R, et al. 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5:** 4767. doi:10.1038/ncomms5767

McCarthy LC. 1996. Whole genome radiation hybrid mapping. *Trends Genet* **12:** 491–493.

McCarthy LC, Bihoreau MT, Kiguwa SL, Browne J, Watanabe TK, Hishigaki H, Tsuji A, Kiel S, Webber C, Davis ME, et al. 2000. A whole-genome radiation hybrid panel and framework map of the rat genome. *Mamm Genome* **11:** 791–795.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47:** D419–D426. doi:10.1093/nar/gky1038

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, et al. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47:** D351–D360. doi:10.1093/nar/gky1100

Morita R, Hirohashi Y, Torigoe T, Ito-Inoda S, Takahashi A, Mariya T, Asanuma H, Tamura Y, Tsukahara T, Kanaseki T, et al. 2016. Olfactory receptor family 7 subfamily C member 1 is a novel marker of colon cancer-initiating cells and is a potent target of immunotherapy. *Clin Cancer Res* **22:** 3298–3309. doi:10.1158/1078-0432.ccr-15-1709

Olarerin-George AO, Hogenesch JB. 2015. Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive. *Nucleic Acids Res* **43:** 2535–2542. doi:10.1093/nar/gkv136

Olivier M, Aggarwal A, Allen J, Almendras AA, Bajorek ES, Beasley EM, Brady SD, Bushard JM, Bustos VI, Chu A, et al. 2001. A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291:** 1298–1302. doi:10.1126/science.1057437

Ouedraogo M, Bettembourg C, Bretaudeau A, Sallou O, Diot C, Demeure O, Lecerf F. 2012. The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. *PLoS One* **7:** e50653. doi:10.1371/journal.pone.0050653

Park CC, Ahn S, Bloom JS, Lin A, Wang RT, Wu T, Sekar A, Khan AH, Farr CJ, Lusis AJ, et al. 2008. Fine mapping of regulatory loci for mammalian gene expression using radiation hybrids. *Nat Genet* **40:** 421–429. doi:10.1038/ng.113

R Core Team. 2018. R: a language and environment for statistical computing.

Reznik E, Miller ML, Şenbabaoğlu Y, Riaz N, Sarungbam J, Tickoo SK, Al-Ahmadie HA, Lee W, Seshan VE, Hakimi AA, et al. 2016. Mitochondrial DNA copy number variation across human cancers. *Elife* **5:** e10769. doi:10.7554/elife.10769

Rupp O, MacDonald ML, Li S, Dhiman H, Polson S, Griep S, Heffner K, Hernandez I, Brinkrolf K, Jadhav V, et al. 2018. A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol Bioeng* **115:** 2087–2100. doi:10.1002/bit.26722

Salzburger W, Steinke D, Braasch I, Meyer A. 2009. Genome desertification in eutherians: can gene deserts explain the uneven distribution of genes in placental mammalian genomes? *J Mol Evol* **69:** 207–216. doi:10.1007/s00239-009-9251-4

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, et al. 2019. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **47:** D23–D28. doi:10.1093/nar/gky1069

Schafer AJ, Farr CJ. 1998. Somatic cell hybrid approaches to genome mapping. In: ICRF Handbook of Genome Analysis, NK Spurr, SP Bryant, BD Young, ed. Vol. 1. (Oxford, UK: Wiley-Blackwell). Chap. 14**,** pp. 321–366.

Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, Murphy TD, Pruitt KD, Thibaud-Nissen F, Albracht D, et al. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* **27:** 849–864. doi:10.1101/gr.213611.116

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47:** D607–D613. doi:10.1093/nar/gky1131

The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47:** D330–D338. doi:10.1093/nar/gky1055

The GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348:** 648–660. doi:10.1126/science.1262110

Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, Hinds DA. 2017. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun* **8:** 599. doi:10.1038/s41467-017-00257-5

Tsai T, Veitinger S, Peek I, Busse D, Eckardt J, Vladimirova D, Jovancevic N, Wojcik S, Gisselmann G, Altmüller J, et al. 2017. Two olfactory receptors-OR2A4/7 and OR51B5-differentially affect epidermal proliferation and differentiation. *Exp Dermatol* **26:** 58–65. doi:10.1111/exd.13132

Walter MA, Spillett DJ, Thomas P, Weissenbach J, Goodfellow PN. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nat Genet* **7:** 22–28. doi:10.1038/ng0594-22

Wang RT, Ahn S, Park CC, Khan AH, Lange K, Smith DJ. 2011. Effects of genome-wide copy number variation on expression in mammalian cells. *BMC Genomics* **12:** 562. doi:10.1186/1471-2164-12-562

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015. Identification and characterization of essential genes in the human genome. *Science* **350:** 1096–1101. doi:10.1126/science.aac7041

Wang X, Kruglyak L. 2014. Genetic basis of haloperidol resistance in *Saccharomyces cerevisiae* is complex and dose dependent. *PLoS Genet* **10:** e1004894. doi:10.1371/journal.pgen.1004894

Wood SN. 2017. Generalized Additive Models: An Introduction with R 2nd ed. (Chapman and Hall/CRC).