Effects of long-term *in vivo* micro-CT imaging on hallmarks of osteopenia and frailty in aging mice

Ariane C. Scheuren[1], Gisela A. Kuhn[1], Ralph Müller[1*]

[1]Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

* Corresponding author

Prof. Ralph Müller, PhD

E: ram@ethz.ch

Abstract

I*n vivo* micro-CT has already been used to monitor microstructural changes of bone in mice of different ages and in models of age-related diseases such as osteoporosis. However, as aging is accompanied by frailty and subsequent increased sensitivity to external stimuli such as handling and anesthesia, the extent to which longitudinal imaging can be applied in aging studies remains unclear. Consequently, the potential of monitoring individual mice during the entire aging process – from healthy to frail status – has not yet been exploited. In this study, we assessed the effects of long-term *in vivo* micro-CT imaging on hallmarks of aging both on a local (i.e., static and dynamic bone morphometry) and systemic (i.e., frailty index (FI) and body weight) level at various stages of the aging process. Furthermore, using a premature aging model ($PolgA^{(D257A/D257A)}$), we assessed whether these effects differ between genotypes.

The 6th caudal vertebrae of 4 groups of mice ($PolgA^{(D257A/D257A)}$ and $PolgA^{(+/+)}$) were monitored by *in vivo* micro-CT every 2 weeks. One group was subjected to 11 scans between weeks 20 and 40 of age, whereas the other groups were subjected to 5 scans between weeks 26-34, 32-40 and 40-46, respectively. The long-term monitoring approach showed small but significant changes in the static bone morphometric parameters compared to the other groups. However, no interaction effect between groups and genotype was found, suggesting that PolgA mutation does not render bone more or less susceptible to long-term micro-CT imaging. The differences between groups observed in the static morphometric parameters were less pronounced in the dynamic morphometric parameters. Moreover, the body weight and FI were not affected by more frequent imaging sessions. Finally, we observed that longitudinal designs including baseline measurements already at young age are more powerful at detecting hallmarks of aging than those including multiple groups with fewer imaging sessions.

Keywords:

Aging, frailty, bone, *in vivo* micro-CT imaging, frailty index

Introduction

With the estimated increase in life expectancy in the next 30 years [1], the number of people suffering from frailty will also substantially increase [2, 3]. Characterized by the decline in multiple physiological functions, which collectively result in the accumulation of health deficits, frailty leads to a higher vulnerability to adverse health outcomes including falls and osteoporotic fractures [2]. Indeed, measuring frailty, using tools such as the frailty phenotype [4] or frailty index (FI) [5], has been shown to be predictive of osteoporotic fractures [6-9]. Hence, the combined assessment of frailty and of skeletal health could be beneficial for the clinical diagnosis of osteoporosis and of frailty.

With the advent of longitudinal *in vivo* phenotyping tools such as the clinical mouse frailty index (FI) [2], animal models of frailty are of increasing interest in aging studies as the accumulation of health deficits can be quantified over time in individual animals [10-12]. Likewise, longitudinal *in vivo* micro-CT imaging – allowing a non-invasive quantitative and qualitative assessment of the 3D bone micro-architecture over time in individual animals – has become of key importance to investigate time-dependent effects of pathologies and/or treatments in preclinical studies [13, 14]. Combined with advanced image registration techniques, bone formation as well as bone resorption activities can be directly quantified by registering consecutive time-lapsed images onto one another [15, 16]. Furthermore, the addition of computational models has been highly valuable to non-invasively estimate bone strength as well as local mechanical properties associated with aging [17, 18] and/or different interventions [19-21]. However, despite the numerous advantages of *in vivo* micro-CT imaging, the effects of cumulative radiation, anesthesia and handling both on a local level –i.e., on the micro-architecture and function of the scanned tissue - as well as on the systemic level – i.e., on the general well-being of the animals - must be considered [13, 14, 22, 23]. Several studies have

investigated the effects of repeated *in vivo* micro-CT imaging on bone morphometric parameters in rodents, however the reports have been controversial [20, 24-30]. While *in vivo* micro-CT imaging has been shown to have dose-dependent effects on bone morphometric parameters in adolescent rats [28], studies in ovariectomized [24, 26] and adult rats [27] have shown no effects of repeated *in vivo* micro-CT imaging. Similarly, studies using mouse models have shown small but significant effects in the trabecular bone compartment [20, 24], whereas other studies reported no imaging associated effects on bone morphometric parameters [25, 29, 30]. Furthermore, the effects of repeated *in vivo* micro-CT imaging on bone morphometry have been shown to be dependent on the age of the animals, with older animals being less sensitive to imaging associated changes in bone morphometry [20]. Conversely, aged mice are known to be more sensitive to handling and anesthesia [31, 32], and hence, the extent to which long-term longitudinal micro-CT imaging can be applied to monitor individual mice during the entire process of aging remains unclear.

Using an established *in vivo* micro-CT approach to monitor the 6^{th} caudal vertebrae [15, 33], we have previously shown that 5 imaging sessions did not have an effect on the bone micro-architecture and remodeling rates in 15-week old C57BL/6 mice. This approach has subsequently been used in middle-aged and aged mice [17]. By combining long-term *in vivo* micro-CT imaging of the 6^{th} caudal vertebrae with longitudinal FI measurements, we have recently identified hallmarks of frailty and senile osteoporosis in the $PolgA^{(D257A/D257A)}$ mutator mouse [34, 35], which, due to a defect in the proofreading activity of its mitochondrial DNA polymerase gamma, exhibits a premature aging phenotype [36, 37]. In this study, we assessed whether this long-term *in vivo* micro-CT imaging approach has biasing effects on the local and systemic level at various stages of the aging process, and whether these effects differ between genotypes. Specifically, we compared static and dynamic bone morphometric parameters as

well as body weight and FI measurements of PolgA$^{(+/+)}$ (in the following referred to as WT) and PolgA$^{(D257A/D257A)}$ (in the following referred to as PolgA) mice subjected to 11 consecutive imaging sessions with those of mice subjected to 5 consecutive imaging sessions at various ages. By performing both cross-sectional comparisons between genotypes and between imaging groups (i.e., that were scanned at different time-points) as well as longitudinal comparisons within individual animals (i.e., that were scanned both at young and old age), we finally aimed to provide important insight for the effective design of studies applying *in vivo* micro-CT imaging in aging mice.

Materials and methods

Study design

To assess potential effects of increased radiation, anesthesia and handling associated with repeated *in vivo* micro-CT imaging, n=88 female mice were aged in parallel and divided into four groups (with n=12 PolgA$^{(D257A/D257A)}$ (referred to as PolgA) and n=10 PolgA$^{(+/+)}$ (referred to as WT) per group). The first group was scanned over 20 weeks (11 scans between the age of 20-40 weeks), whereas the other groups were scanned over 8 weeks (5 scans between weeks 26-34, 32-40 and 40-46, respectively) as illustrated in Fig 1. For each group, the two last scans overlapped with the two first scans of the subsequent group to allow comparison of the dynamic morphometric parameters between groups. All mouse experiments described in the present study were carried out in strict accordance with the recommendations and regulations in the Animal Welfare Ordinance (TSchV 455.1) of the Swiss Federal Food Safety and Veterinary Office and were approved by the local authorities (license numbers 262/2016, Verterinäramt des Kantons Zürich, Zurich, Switzerland).

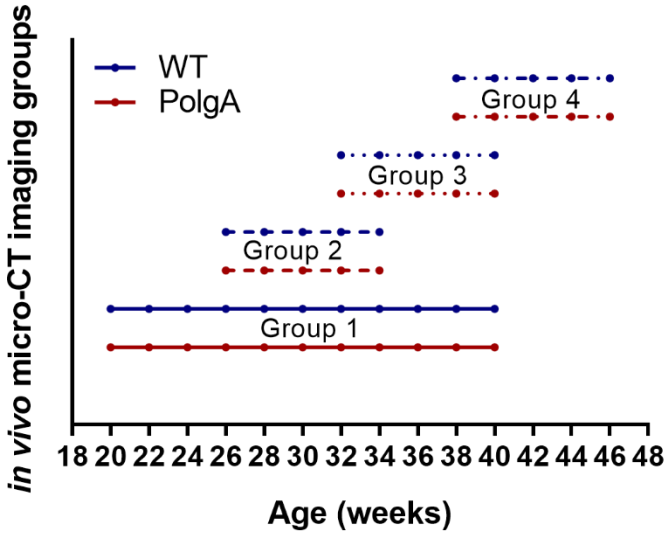


Fig 1 Illustration of the study design showing the time-points and duration of *in vivo* micro-CT imaging for the four groups of mice.

Each group consists of n=12 PolgA (displayed in red) and n=10 WT (displayed in blue).

Animals

A colony of the mouse strain expressing an exonuclease-deficient version of the mitochondrial DNA polymerase γ (PolgA$^{(D257A)}$, B6.129S7(Cg)-Polg^{tm1Prol}/J, JAX stock 017341, The Jackson Laboratory, Farmington CT, USA) was bred and maintained at the ETH Phenomics Center (12h:12h light-dark cycle, maintenance feed and water ad libitum, 3-5 animals/cage) as previously described [34, 35]. The presence of the PolgA knock-in mutation was confirmed by extracting DNA from ear clips (Sigma-Aldrich, KAPA Express Extract, KK7103) followed by qPCR (Bio-Rad, SsoAdvanced Universal SYBR Green Supermix, 1725272) and melt curve analysis. The primers used for genotyping (5' to 3'; Rev Common: AGT AGT CCT GCG CCA ACA CAG; Wild type forward: GCT TTG CTT GAT CTC TGC TC; Mutant forward: ACG AAG TTA TTA GGT CCC TCG AC) were recommended by the Jackson Laboratory.

Micro-CT imaging and analysis

In vivo micro-CT (vivaCT 40, Scanco Medical AG, isotropic nominal resolution: 10.5 µm; 55 kVp, 145 µA, 350 ms integration time, 500 projections per 180°, scan duration ca. 15 min, radiation dose per scan ca. 640 mGy) scans of the 6th caudal vertebrae were performed every 2 weeks. Animals were anesthetized with isoflurane (induction/maintenance: 5%/1-2% isoflurane/oxygen). Micro-CT data was processed and standard bone microstructural parameters were calculated in trabecular, cortical and whole bone by using automatically selected masks for these regions as described previously [33]. To calculate dynamic morphometric parameters, micro-CT images from consecutive time-points were registered onto one another. The voxels present only at the initial time point were considered resorbed whereas voxels present only at the later time point were considered formed. Voxels that were present at both time points were considered as quiescent bone. By overlaying the images, morphometrical analysis of bone formation and resorption sites within the trabecular region allowed calculations of bone formation rate (BFR), bone resorption rate (BRR), mineral apposition rate (MAR), mineral resorption rate (MRR), mineralizing surface (MS) and eroded surface (ES) [15].

Quantification of the clinical mouse frailty index (FI)

As recommended in the recently established toolbox for the longitudinal assessment of healthspan in aging mice [38], frailty was quantified using the clinical mouse FI [39], which includes the assessment of 31 non-invasive clinical items. For 29 of these items, mice were given a score of 0 if not present, 0.5 if there was a mild deficit, and 1 for a severe deficit. The final two items were weight and body surface temperature, which were scored based on the number of standard deviations from a reference mean in young adult mice as previously described [39].

Statistical analysis

Data are represented as mean±SD. For analysis of the longitudinal micro-CT images, frailty index and body weight measurements, linear mixed model analysis was performed using the lmerTEST package [40] in R (R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). Fixed effects were allocated to age, genotype and group and a random effect was allocated to the individual mice to account for inherent variability between mice. Furthermore, an interaction effect between age and genotype as well as between genotype and groups were assessed. For the comparison of the bone morphometric parameters, frailty index and body weights at 40 weeks of age, values of the individual mice are shown. Effects of genotype and imaging group were analyzed via two-way ANOVA followed by Tukey's multiple comparison test or one-way ANOVA, respectively using SPSS (IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY, USA). Power analysis was performed in G*Power (G*Power, Version 3.1.3., Düsseldorf, Germany [41]) and in R (R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).

Results

Effect of genotype on bone morphometry and frailty

Fig 2 shows the changes in static bone morphometric parameters over time in the different imaging groups obtained by longitudinal *in vivo* micro-CT imaging. Taking all the longitudinal micro-CT measurements of the different groups into account, PolgA had lower bone volume fraction (BV/TV, -10%, $p<0.0001$), trabecular thickness (Tb.Th, -6%, $p<0.0001$), cortical area fraction (Ct.Ar/Tt.Ar, -6%, $p<0.0001$) and cortical thickness (Ct.Th, -5%, $p<0.0001$) compared

to WT (Fig 2A-D). The trabecular number (Tb.N) and trabecular spacing (Tb.Sp) were not significantly different between genotypes ($p>0.05$). Furthermore, the age-related changes in bone morphometric parameters developed differently between genotypes with a significant interaction effect between age and genotype for BV/TV, Tb.Th, Ct.Ar/Tt.Ar and Ct.Th ($p<0.0001$, Fig 2). While BV/TV, trabecular thickness (Tb.Th), Ct.Ar/Tt.Ar and cortical thickness (Ct.Th) initially increased in both genotypes, the increase in these parameters ceased to continue in PolgA mice from 30-32 weeks onwards (Fig 2A-D).

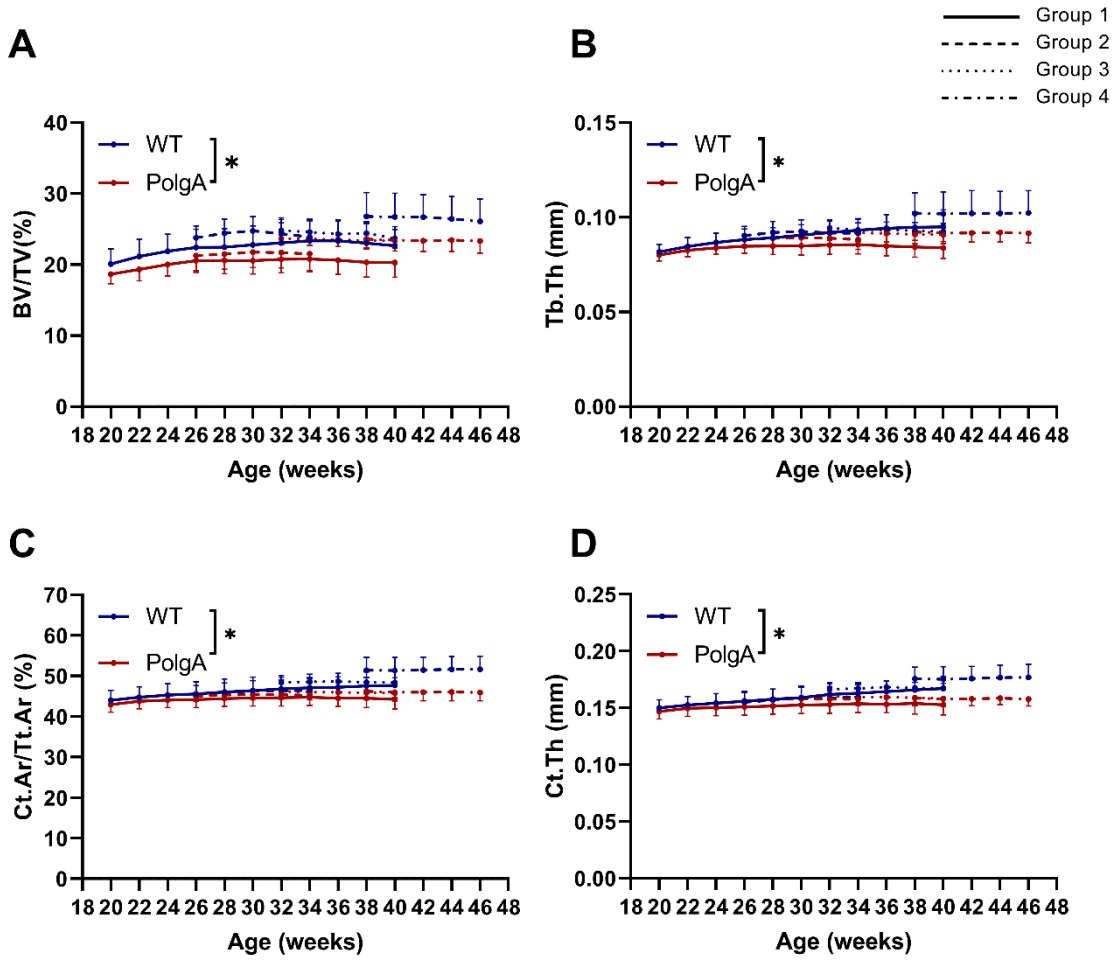
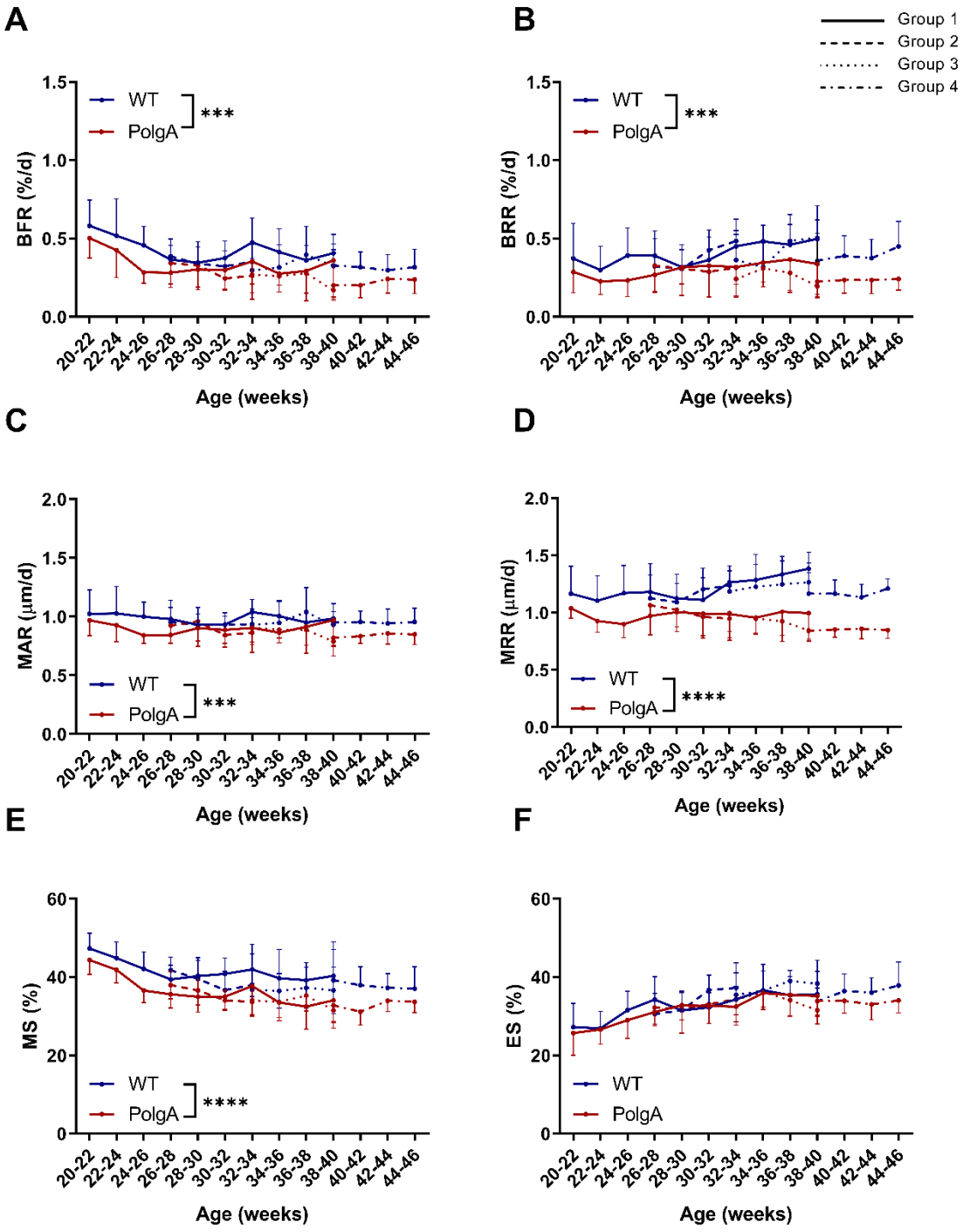


Fig 2 Static bone morphometric parameters obtained by longitudinal *in vivo* micro-CT monitoring of the 6th caudal vertebrae between 20 and 46 weeks of age.

(A) Bone volume fraction (BV/TV), (B) trabecular thickness (Tb.Th), (C) cortical area fraction (Ct.Ar/Tt.Ar) and (D) cortical thickness (Ct.Th). (* $p<0.05$ PolgA (red lines) vs WT (blue lines)

determined by linear mixed model. The different patterns represent different groups of mice scanned at different time-points)

By registering consecutive time-lapsed *in vivo* micro-CT images onto one-another [15], we furthermore assessed the dynamic remodeling activities in PolgA and WT mice. On average, PolgA mice had significantly lower bone formation rate (BFR, -27%, p=0.001) and bone resorption rate (BRR, -24%, p=0.001) compared to WT (Fig 3A,B). BFR did not develop differently between genotypes, whereas there was a significant interaction effect between age and genotype for BRR (p=0.014). Mineral apposition (MAR) and resorption rate (MRR), which represent the thickness of formation and resorption packages, were lower (-9%, p=0.0013 and -18%, p<0.0001) in PolgA mice compared to WT (Fig 3C,D). While MAR did not develop differently between genotypes, MRR increased in WT but remained constant in PolgA mice, thus leading to a significant interaction effect between age and genotype for MRR (p<0.001). The mineralized surface (MS), which represents the surfaces of formation sites was lower (-13%, p<0.0001) in PolgA mice compared to WT whereas ES was similar between genotypes (p>0.05, Fig 3E,F). Neither MS nor ES showed an interaction effect between age and genotype (p>0.05). Overall, these results suggest that PolgA mice have lower bone remodeling activities compared to WT.



198

199 **Fig 3 Dynamic bone morphometric parameters obtained by longitudinal *in vivo***
200 **monitoring of the 6th caudal vertebrae between 20 and 46 weeks of age.**

201 (A) Bone formation rate (BFR), (B) bone resorption rate (BRR), (C) mineral apposition rate
202 (MAR), (D) mineral resorption rate (MRR), (E) mineralizing surface (MS) and (F) eroded
203 surface (ES). (* $p<0.05$ PolgA (red lines) vs WT (blue lines) determined by linear mixed model.
204 The different line patterns represent different groups of mice scanned at different time-points)

Fig 4 shows the changes in clinical mouse frailty index (FI) and body weights of PolgA and WT mice in the different imaging groups. In line with the known accelerated aging phenotype of PolgA mice [36, 37], longitudinal assessments of the FI showed that the mean FI averaged over all time-points was significantly higher in PolgA (+98%, $p<0.0001$) compared to WT (Fig 4A). Furthermore, the FI developed differently with age in PolgA mice compared to WT (interaction effect between age and genotype, $p<0.0001$, Fig 4A). While PolgA and WT mice had similar FI scores at 34 weeks, PolgA mice continuously accumulated health deficits (i.e., graying, ruffled fur, distended abdomen) with age leading to higher FI scores compared to WT from 38 weeks onwards. Conversely, the body weight continuously increased in both genotypes, with no differences detected between genotypes ($p>0.05$, Fig 4B).

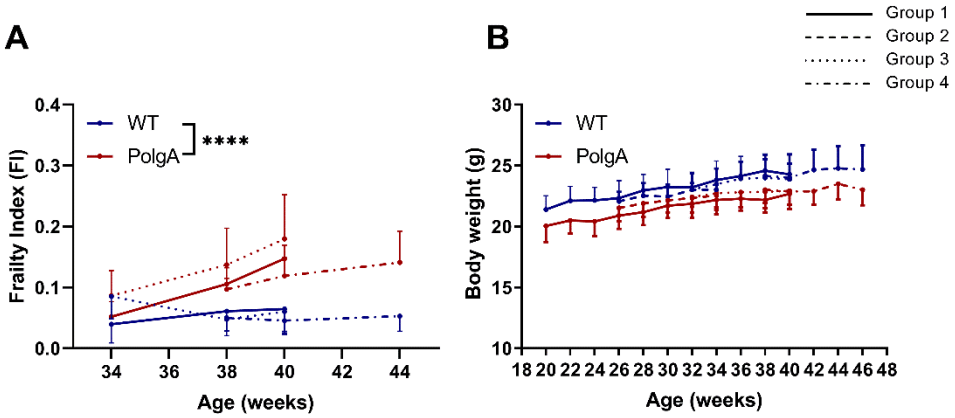


Fig 4 Longitudinal quantification of the (A) frailty index (FI) and (B) body weight in PolgA and WT mice between 20 and 46 weeks of age.

(**** $p<0.0001$ PolgA (red lines) vs WT (blue lines) determined by linear mixed model. The different line patterns represent different groups of mice scanned at different time-points).

Effect of *in vivo* micro-CT imaging on bone morphometry and frailty

In order to address whether the cumulative effects of radiation, anesthesia and handling associated with long-term *in vivo* micro-CT imaging influenced the bone phenotype of PolgA and WT mice, linear regression analysis was performed on the static and dynamic parameters of all four groups (see study design illustrated in Fig 1). As is already visually evident in Fig 2,

there was a considerable heterogeneity in static bone morphometric parameters between the different groups. Compared to the first group which was scanned over 20 weeks, the fourth group showed higher BV/TV ($p<0.001$) and Tb.Th ($p=0.003$, Fig 2A,B), along with higher Tb.N ($p=0.002$) and lower Tb.Sp ($p<0.0001$). With respect to the cortical bone, the fourth group also showed higher Ct.Ar/Tt.Ar ($p<0.0001$) and Ct.Th ($p=0.005$) compared to the first group. However, none of the morphometric parameters showed a significant interaction effect between genotype and group.

The differences between groups observed in the dynamic morphometric parameters were less pronounced than in the static morphometric parameters. Averaged over all time-points, a significant effect of group was only found for BFR ($p=0.007$), BRR ($p=0.01$) and MRR ($p=0.01$, Fig 3A,B,D). Post-hoc analysis between groups showed that WT mice of the fourth group had significantly lower BRR ($p=0.029$) and MRR ($p=0.025$) compared to WT of the first group, thus explaining the higher BV/TV and Tb.Th observed in that group. Similar to the static morphometric parameters, no significant interaction effects between genotype and age were found for any of the parameters. Linear regression analysis of the longitudinal measurements of FI and body weights did not show any differences between imaging groups ($p>0.05$, Fig 4A,B).

Effect of imaging session number on bone morphometry and frailty

In addition to evaluating the longitudinal data, the bone morphometric parameters and FI of the different imaging groups were cross-sectionally compared at 40 weeks of age, i.e. the time-point at which the different groups had received either 11 scans, 5 scans or 2 scans, respectively (Figs 5 and 6). Both genotype and imaging group significantly affected BV/TV, Tb.Th, Ct.Ar/Tt.Ar and Ct.Th (Fig 5A-D and Table S1 for p-values and effect sizes *f*), however none of the parameters showed a significant interaction effect between genotype and group. On average, PolgA mice had lower BV/TV (-8.2%, $p=0.001$), Tb.Th (-7.9%, $p<0.0001$), Ct.Ar/Tt.Ar (-7.8%,

p<0.0001) and Ct.Th (-8%, p<0.0001) compared to WT. Furthermore, for Tb.Th, Ct.Ar/Tt.Ar and Ct.Th, the effect of genotype was 1.3, 1.75 and 2.5 fold stronger than the effect of imaging session number (group) as shown by the higher effect sizes *f* determined by two-way ANOVA analysis (Table S1). Conversely, for BV/TV, the effect of imaging session number was 1.48 fold stronger compared to the effect of genotype. Post-hoc analysis within genotypes showed that WT mice of the 11-scan group had lower BV/TV (p=0.008) and Ct.Ar/Tt.Ar (p=0.032) compared to those of the 2-scan group (Fig 5A,C). Similar to WT mice, PolgA of the 11-scan group showed lower BV/TV (p<0.0001) and lower Tb.Th (p=0.005 and p=0.002) compared to the 5- and 2-scan groups, with no differences between imaging groups detected in the cortical morphometric parameters (Fig 5B-D).

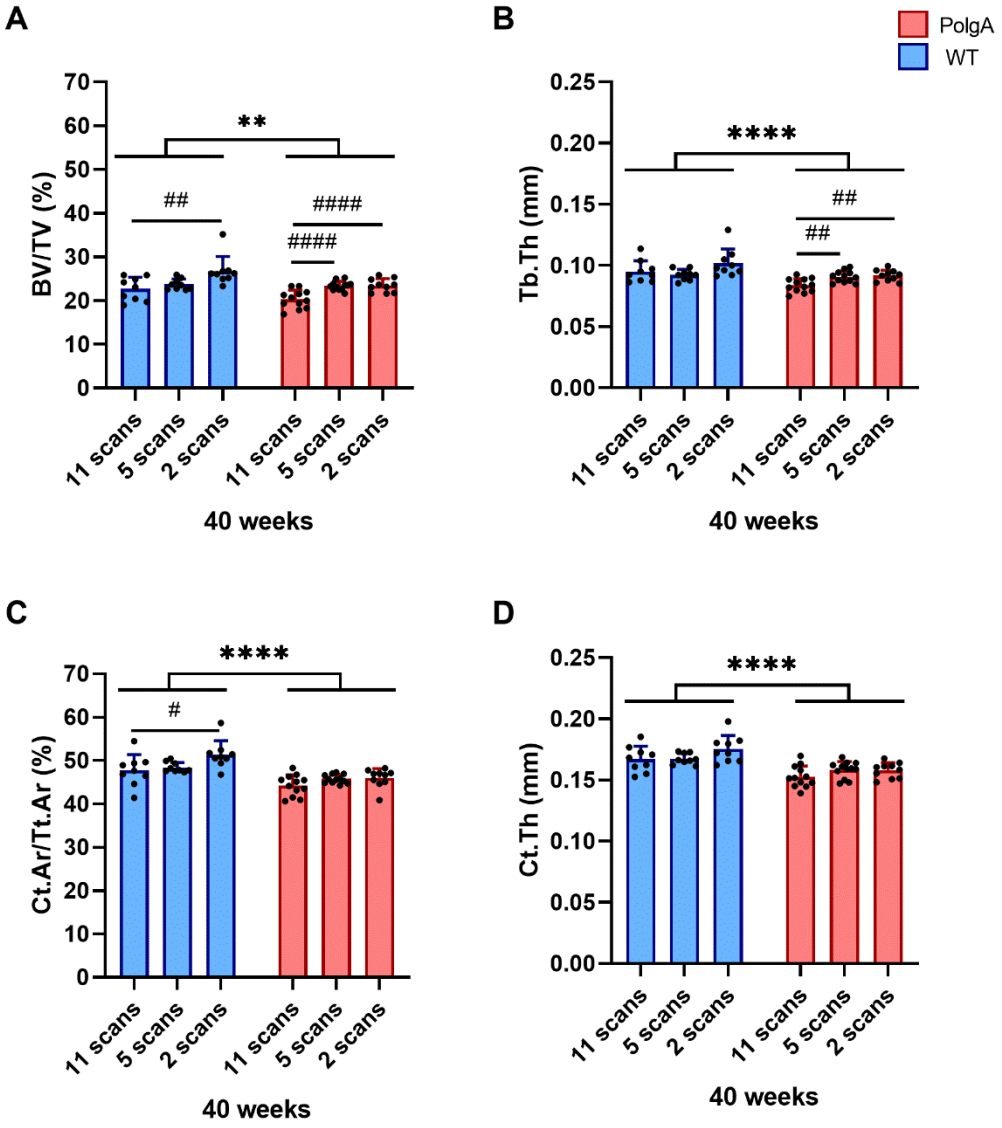


Fig 5 Static bone morphometric parameters at 40 weeks of age for the different groups having received either 11 scans, 5 scans or 2 scans, respectively.

(A) bone volume fraction (BV/TV), (B) trabecular thickness (Tb.Th), (C) cortical area fraction (Ct.Ar/Tt.Ar), (D) cortical thickness (Ct.Th). ** $p<0.01$ and **** $p<0.0001$ between genotypes determined by two-way ANOVA, # $p<0.05$, ## $p<0.01$ and #### $p<0.0001$ post-hoc test within genotypes determined by one-way ANOVA and Tukey's multiple comparisons test.

With respect to the dynamic parameters, both genotype and imaging session number significantly affected BFR, BRR and MRR (Fig 6A,B,D and Table S1 for p-values and effect sizes *f*), however none of the parameters showed a significant interaction effect between genotype and imaging session number. On average, PolgA mice had lower BFR (-30.8%,

p=0.005), BRR (-41.9%, p<0.0001), MAR (-10%, p=0.029 and MRR (-29.7%, p<0.0001) compared to WT. Furthermore, for BRR and MRR, the effect of genotype was 1.3 and 2.4 fold stronger than the effect of imaging session number as shown by the higher effect sizes *f* determined by two-way ANOVA analysis (Table S1). Conversely, for BFR and MAR, the effect of imaging session number was 1.1 fold stronger compared to the effect of genotype. Furthermore, WT mice of the 11-scan group had higher MRR compared to WT mice of the 2-scan group (p=0.026, Fig 6H). PolgA mice of the 11-scan group had higher BFR (p=0.019) and BRR (p=0.022) compared to PolgA mice of the 5-scan group (Fig 6E,F), whereas no differences were observed between 11- and 2-scan group, respectively.

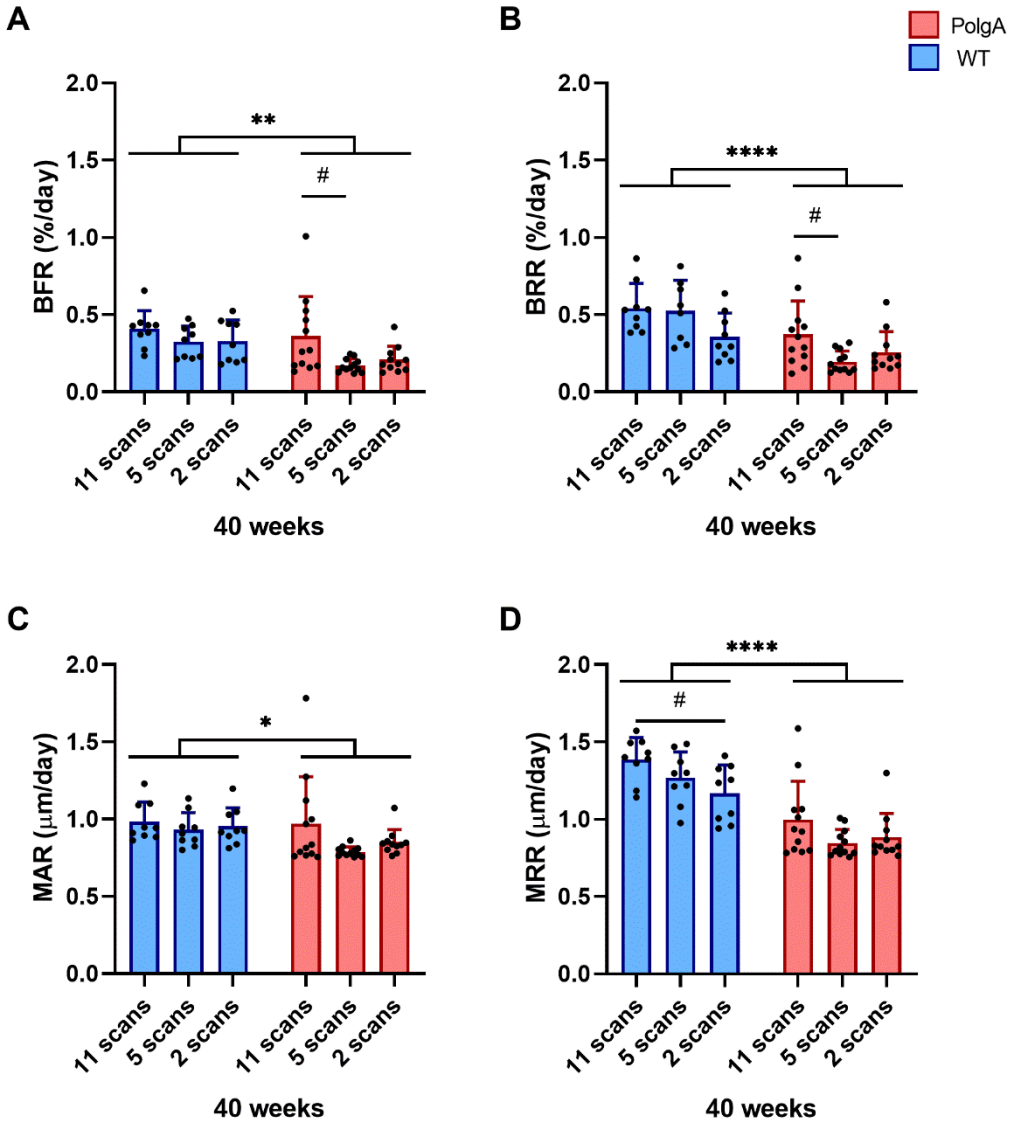


Fig 6 Dynamic bone morphometric parameters at 40 weeks of age for the different groups having received either 11 scans, 5 scans or 2 scans, respectively.

(A) Bone formation rate (BFR), (B) bone resorption rate (BRR), (C) mineral apposition rate, (D) mineral resorption rate (MRR). * $p<0.05$, ** $p<0.01$ and **** $p<0.0001$ between genotypes determined by two-way ANOVA, # $p<0.05$, post-hoc test within genotypes determined by one-way ANOVA and Tukey's multiple comparisons test.

At 40 weeks of age, PolgA mice had significantly higher FI compared to WT (+ 166%, $p<0.0001$, Fig 7A). The comparison between imaging groups showed that PolgA mice in the 5-scan group had significantly higher FI scores compared to PolgA mice of the 2-scan group

(p=0.019). No significant differences between genotypes or scanning groups were detected for the body weight at 40 weeks of age (Fig 7B).

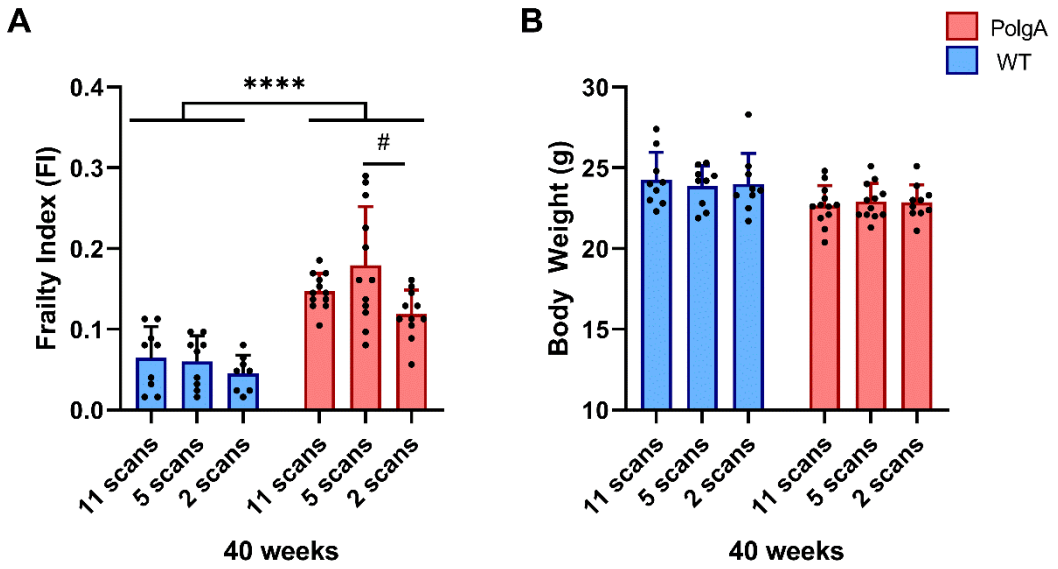


Fig 7 Comparison of (A) FI and (B) body weights at 40 weeks of age in PolgA (red bars) and WT (blue bars) mice. (****p<0.0001 between genotypes and #p<0.05 determined by post-hoc one-way ANOVA and Tukey's multiple comparisons test).

Considering the small sample sizes (n=9 per group in WT and n=12 per group in PolgA) used for this cross-sectional comparison between scanning groups, the achieved power of the analysis was computed given the obtained effect sizes (*f*) and alpha level of 0.05 (Table S2). While sufficient power (≥ 0.8) was achieved to detect any effects that might have existed in the trabecular morphometric parameters, the power for detecting differences in the cortical as well as dynamic morphometric parameters was not sufficient, suggesting that a higher number of samples would be required to detect any differences between scanning groups. Conversely, by performing a longitudinal comparison (paired t-test between parameters of individual WT and PolgA mice measured both at 20 and 40 weeks of age), sufficient power (power ≥ 0.8) was obtained for all static morphometric parameters. Nevertheless, as the dynamic bone morphometric parameters in PolgA mice remained relatively constant between 20 and 40 weeks

of age, the effect sizes were small and hence, higher samples sizes would be beneficial to detect differences in dynamic bone morphometric parameters over time in individual mice.

Discussion

Inspired by a previous study, which characterized the bone phenotype and the development of frailty in the PolgA mouse model [34, 35], this study assessed the impact of long-term *in vivo* micro-CT imaging on hallmarks of osteopenia and frailty in individual mice during the process of aging. The unique study design, for which a large cohort of animals was aged in parallel up to 46 weeks of age, provided the possibility not only for cross-sectional comparisons between genotypes and between imaging groups (i.e., that were scanned at different time-points) but also for longitudinal comparisons within individual animals (i.e., that were scanned both at young and old age).

In agreement with previous studies of cross-sectional [36, 37, 42] and longitudinal designs [34, 35], PolgA and WT had similar bone morphometric parameters at 20 weeks of age, which then diverged over time such that PolgA had significantly lower bone volume and quality at 40 weeks of age. Concomitantly, PolgA accumulated multiple health deficits over time (e.g., graying, ruffled fur, distended abdomen, kyphosis) leading to a significantly higher FI in PolgA mice from 38 weeks onwards. The clear difference in the bone morphometric parameters and FI between genotypes was observed both when groups were cross-sectionally compared and when individual mice were monitored over time. Interestingly though, *in vivo* micro-CT imaging over 20 weeks showed that this difference in bone micro-architecture was not due to bone loss in PolgA mice, but rather in the inability to reach peak bone mass, of which a more comprehensive description has been provided elsewhere [34, 35]. Interestingly, the registration of consecutive micro-CT images revealed that PolgA mice had lower bone remodeling activities compared to

WT as shown by reduced bone formation and resorption rates, with no differences in the net remodeling rate. Similarly, senile osteoporosis in humans is characterized by low bone turnover, as opposed to the high bone turnover rates (higher resorption activities) observed during postmenopausal osteoporosis [43-45].

The comparison between different imaging groups revealed a considerable heterogeneity between groups with the fourth group having higher trabecular and cortical bone morphometric parameters compared to the other groups. However, we did not observe an interaction effect between genotype and imaging groups, suggesting that the PolgA mutation does not render bone more or less susceptible to cumulative effects of radiation, anesthesia and handling associated with *in vivo* micro-CT imaging. Hence, the comparison between two genotypes remains valid despite the potential confounding effects of more frequent *in vivo* micro-CT imaging sessions. Furthermore, with the exception of BV/TV, the effect of genotype on static bone morphometric parameters was larger than the effect of repeated *in vivo* micro-CT imaging. That *in vivo* micro-CT imaging associated effects on bone morphometric parameters are stronger in trabecular bone compared to cortical bone has previously been shown [24]. In that study, the tibiae of sham- and ovariectomized-C57Bl/6J mice showed lower trabecular bone volume compared to the contralateral non-irradiated limbs, whereas this effect was not observed in healthy untreated 8- to 10-week-old control C57Bl/6J mice [24]. In line with our study, no interaction effect between ovariectomy and *in vivo* micro-CT imaging was found. This suggests that despite small but significant effects of *in vivo* micro-CT imaging on bone morphometric parameters, the comparison between different study groups/treatments remains valid. Willie et al. have also reported lower BV/TV in 10-week-old C57BL/6J mice subjected to multiple *in vivo* micro-CT scans compared to age-matched mice subjected to only one *in vivo* micro-CT scan [20]; however, this effect was not observed in 26-week old mice, suggesting that imaging

356 associated effects are stronger in younger mice. Conversely, other studies have reported no
357 imaging associated effects on bone morphometric parameters in mice ranging from pre-pubertal
358 to adult age [25] up to late adulthood (48 weeks) [29]. Using the same micro-CT settings as in
359 the present study, we have previously shown that five scans did not have an effect on the bone
360 microstructure or bone remodeling rates in the caudal vertebrae of 15-week-old C57BL/6 mice
361 [17, 33]. More recently, we have furthermore used a similar *in vivo* micro-CT approach to
362 monitor specific healing phases after osteotomy and did not observe any significant imaging-
363 associated changes in bone volume and turnover in the fracture callus of mice after 7 scans [30].
364 In the present study, the comparison between groups at 40 weeks of age showed that the PolgA
365 mice that were subjected to 11 imaging sessions had lower trabecular bone morphometric
366 parameters compared to those subjected to 5 and 2 sessions, respectively, while no differences
367 between any of the groups were detected in the cortical bone. For the WT mice, the group
368 subjected to 2 imaging sessions showed higher BV/TV and Ct.Ar/Tt.Ar compared to those
369 subjected to 11 sessions, while no significant differences between 11 and 5 imaging sessions
370 were found. Hence, although effects associated with multiple time-lapsed micro-CT scans seem
371 to be present when compared to a very low number of 2 imaging sessions, there does not seem
372 to be major differences between performing 5 or 11 sessions. Furthermore, although small but
373 significant effects were observed between imaging groups in the static bone morphometric
374 parameters, the differences between groups were less pronounced in the dynamic remodeling
375 parameters. Taking all the longitudinal data into account, no differences between groups were
376 observed for the parameters associated with bone formation. For parameters associated with
377 bone resorption, the WT mice of the first group showed higher BRR and MRR compared to the
378 fourth group. When the groups were compared at 40 weeks of age (having received either 11
379 or 2 imaging sessions, respectively), the 11-scan group showed higher MRR compared to the

2-scan group, suggesting a potential imaging-induced increase in osteoclast activity. An increased number and activity of osteoclasts has previously been reported in C57Bl/6 mice subjected to whole-body irradiation with an X-ray dosage of 2Gy [46], however, as *in vivo* micro-CT imaging is localized to the scanned area, these results are not directly comparable to whole-body irradiation. One study using *in vivo* micro-CT showed dose-dependent effects on bone formation and resorption activities; while 3 consecutive scans at a high dose (776 mGy) resulted in increased bone resorption but no differences in bone formation in the tibiae of 10-week-old C57Bl/6J mice, no effects were observed when scanning at a lower dose (434mGy) [25]. In line with these results obtained *in vivo*, scanning at the lower dose did not affect the osteogenic differentiation of bone marrow osteoprogenitors as well as osteoclast formation from bone marrow osteoclast precursors *in vitro* [25]. Although the dose used in the current study (640 mGy) lies in between the two doses reported previously, caudal vertebrae – which predominantly contain yellow (fatty) bone marrow – should be less sensitive to radiation compared to long bones, which predominantly contain red bone marrow [47, 48]. Nevertheless, owing to the low statistical power achieved in the cross-sectional comparison between imaging groups, further studies with higher animal numbers would be necessary to detect cumulative effects of long-term *in vivo* micro-CT imaging on dynamic bone morphometric parameters.

There are several limitations to this study. Firstly, we do not have baseline measurements of all mice at 20 weeks of age, making it impossible to know whether the differences observed between imaging groups were due to repeated micro-CT imaging or due to initial variation between animals. In this respect, we observed that longitudinal designs including baseline measurements already at young age are more powerful at detecting age-related phenotypic changes compared to those including multiple groups with fewer imaging sessions. Hence, although sample sizes of 10-12 animals per group are sufficient for longitudinal studies, higher

404 animal numbers would be beneficial for cross-sectional comparisons of aging mice. A second
405 limitation of this study was the inability to single out the effects of radiation, anesthesia and
406 handling, respectively as the main factor influencing bone morphometric parameters, as more
407 frequent imaging sessions comprise more frequent anesthesia and handling as well as more
408 frequent fixation in the mouse holder used for imaging. Therefore, in order to rule out potential
409 harmful effects of the imaging procedure, future studies should include a control group,
410 receiving a baseline and end-point *in vivo* scan as well as sham-scans for the rest of the
411 measurements. Indeed, a previous study, which included sham-scanning did not observe any
412 imaging associated effects on the bone micro-architecture and cell viability of bone marrow
413 cells in aged rat tibiae subjected to 8 weekly scans [27]. Lastly, although this study assessed the
414 effects of long-term *in vivo* micro-CT imaging on bone remodeling activities, thereby providing
415 indirect information on cellular activities and recruitment, potential damaging effects of this
416 longitudinal micro-CT imaging approach were not assessed at the cellular level. Future studies
417 should therefore include analyses of gene and protein expression in response to long-term *in*
418 *vivo* micro-CT imaging.

419 Interestingly, the heterogeneity in bone morphometry observed between imaging groups was
420 not present for the FI measurements. At 40 weeks of age, the PolgA mice of the 5-scan group
421 had significantly higher FI scores compared to those of the 2 scan group; however, as no
422 differences were detected between the 11- and 2-scan groups, we expect that this difference
423 was related to variation between animals rather than to the cumulative effects of radiation and
424 anesthesia associated with *in vivo* micro-CT imaging. The imaging groups did not show any
425 differences in body weights, which increased over time both in PolgA and WT mice. The
426 absence of weight loss further supports the fact that the well-being of the animals was not
427 negatively affected by the higher number of *in vivo* micro-CT scans in the first group, compared

to groups 2,3 and 4, respectively. These results demonstrate how the parallel tracking of the FI as an addition to longitudinal *in vivo* micro-CT imaging not only allows to link age-related changes in bone morphometry to the development of frailty but also provides a useful tool to assess whether treatments/interventions have biasing effects on the overall health status of the animals. By maximizing the data obtained from individual animals, the total number of animals can be reduced. Based on the results of this study, we therefore recommend that in addition to acquiring baseline measurements of body weight and of the bone micro-architecture, baseline FI measurements should be included for studies in aging mice. Stratifying the animals according to these initial measurements could also be highly valuable to reduce the variability between study groups.

In conclusion, the combination of the longitudinal assessments of the FI and time-lapsed *in vivo* micro-CT imaging allowed for the detection of hallmarks of osteopenia and aging across multiple systems. Although the long-term monitoring approach can potentially lead to small but significant changes in bone morphometric parameters, the comparison between genotypes was not impaired. Moreover, more frequent *in vivo* micro-CT imaging did not negatively affect the multi-system hallmarks of aging such as body weight and frailty index. In line with the goal of "Reduction", the second principle of the 3R's, long-term *in vivo* micro-CT imaging allows to reduce the number of animals required for experiments while maintaining sufficient statistical power to reach a valid conclusion and thus, provides a powerful tool for usage in aging studies.

Acknowledgements

The authors gratefully acknowledge valuable inputs from Dr. Ilaria Bellantuono concerning the establishment of the frailty index at the institute for biomechanics. This manuscript is based upon work supported by the European Cooperation in Science and Technology (COST Action

451 BM1402: MouseAGE) and the European Research Council (ERC Advanced MechAGE ERC-
452 2016-ADG-741883).

453 Conflict of Interest

454 The authors declare no conflict of interest.

455

References

[1] Nations U, *World Population Ageing 2019 (ST/ESA/SER.A/444)*, Department of Economic and Social Affairs, Population Division. www.unpopulation.org, 2020: New York, USA.

[2] Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet (London, England), 2013; 381(9868):752-762.

[3] I. Bellantuono RD, D. Ehninger, A. Fernandes, S. E. Howlett, R. Müller, P. Potter, T. Tchkonia, A.-U. Trendelenburg, J. L. Trejo, R. Vandenbroucke, R. van Os, N. van Riel. Find drugs that delay many diseases of old age. Nature, 2018; 554:293-295.

[4] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA, Collabor CHS. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci, 2001; 56(3):M146-M156.

[5] Jones DM, Song XW, Rockwood K. Operationalizing a frailty index from a standardized comprehensive geriatric assessment. J Am Geriatr Soc, 2004; 52(11):1929-1933.

[6] Kennedy CC, Ioannidis G, Rockwood K, Thabane L, Adachi JD, Kirkland S, Pickard LE, Papaioannou A. A Frailty Index predicts 10-year fracture risk in adults age 25 years and older: results from the Canadian Multicentre Osteoporosis Study (CaMos). Osteoporos Int, 2014; 25(12):2825-2832.

[7] Fang X, Shi J, Song X, Mitnitski A, Tang Z, Wang C, Yu P, Rockwood K. Frailty in relation to the risk of falls, fractures, and mortality in older Chinese adults: Results from the Beijing longitudinal study of aging. J Nutr Health Aging, 2012; 16(10):903-907.

[8] Ensrud KE, Ewing SK, Taylor BC, Fink HA, Stone KL, Cauley JA, Tracy JK, Hochberg MC, Rodondi N, Cawthon PM, for the Study of Osteoporotic Fractures Research G.

Frailty and risk of falls, fracture, and mortality in older women: the study of osteoporotic fractures. J Gerontol A Biol Sci Med Sci, 2007; 62(7):744-751.

[9] Li G, Thabane L, Papaioannou A, Ioannidis G, Levine MAH, Adachi JD. An overview of osteoporosis and frailty in the elderly. BMC Musculoskel Dis, 2017; 18(1):46.

[10] Kane AE, Hilmer SN, Mach J, Mitchell SJ, de Cabo R, Howlett SE. Animal models of frailty: current applications in clinical research. Clin Interv Aging, 2016; 11:1519-1529.

[11] von Zglinicki T, Varela-Nieto I, Brites D, Karagianni N, Ortolano S, Georgopoulos S, Cardoso AL, Novella S, Lepperdinger G, Trendelenburg A-U, van Os R. Frailty in mouse ageing: A conceptual approach. Mech Ageing Dev, 2016; 160(Supplement C):34-40.

[12] Banga S, Heinze-Milne SD, Howlett SE. Rodent models of frailty and their application in preclinical research. Mech Ageing Dev, 2019; 179:1-10.

[13] Tremoleda JL, Khalil M, Gompels LL, Wylezinska-Arridge M, Vincent T, Gsell W. Imaging technologies for preclinical models of bone and joint disorders. EJNMMI Res, 2011; 1(1):11-11.

[14] Dall'Ara E, Boudiffa M, Taylor C, Schug D, Fiegle E, Kennerley AJ, Damianou C, Tozer GM, Kiessling F, Müller R. Longitudinal imaging of the ageing mouse. Mech Ageing Dev, 2016; 160:93-116.

[15] Schulte FA, Lambers FM, Kuhn G, Müller R. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. Bone, 2011; 48(3):433-442.

[16] Birkhold AI, Razi H, Duda GN, Weinkamer R, Checa S, Willie BM. Mineralizing surface is the main target of mechanical stimulation independent of age: 3D dynamic in vivo morphometry. Bone, 2014; 66:15-25.

[17] Lambers FM, Kuhn G, Weigt C, Koch KM, Schulte FA, Müller R. Bone adaptation to cyclic loading in murine caudal vertebrae is maintained with age and directly correlated to the local micromechanical environment. J Biomech, 2015; 48(6):1179-1187.

[18] Razi H, Birkhold AI, Zaslansky P, Weinkamer R, Duda GN, Willie BM, Checa S. Skeletal maturity leads to a reduction in the strain magnitudes induced within the bone: a murine tibia study. Acta Biomaterialia, 2015; 13:301-310.

[19] Levchuk A, Zwahlen A, Weigt C, Lambers FM, Badilatti SD, Schulte FA, Kuhn G, Muller R. The Clinical Biomechanics Award 2012 - presented by the European Society of Biomechanics: large scale simulations of trabecular bone adaptation to loading and treatment. Clin Biomech (Bristol, Avon), 2014; 29(4):355-362.

[20] Willie BM, Birkhold AI, Razi H, Thiele T, Aido M, Kruck B, Schill A, Checa S, Main RP, Duda GN. Diminished response to in vivo mechanical loading in trabecular and not cortical bone in adulthood of female C57Bl/6 mice coincides with a reduction in deformation to load. Bone, 2013; 55(2):335-346.

[21] Yang H, Butz KD, Duffy D, Niebur GL, Nauman EA, Main RP. Characterization of cancellous and cortical bone strain in the in vivo mouse tibial loading model using microCT-based finite element analysis. Bone, 2014; 66:131-139.

[22] Hildebrandt IJ, Su H, Weber WA. Anesthesia and other considerations for in vivo imaging of small animals. ILAR Journal, 2008; 49(1):17-26.

[23] Hohlbaum K, Bert B, Dietze S, Palme R, Fink H, Thöne-Reineke C. Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice - Assessing the degree of distress. PloS One, 2017; 12(6):e0179588.

[24] Klinck RJ, Campbell GM, Boyd SK. Radiation effects on bone architecture in mice and rats resulting from in vivo micro-computed tomography scanning. Med Eng Phys, 2008; 30(7):888-895.

[25] Laperre K, Depypere M, van Gastel N, Torrekens S, Moermans K, Bogaerts R, Maes F, Carmeliet G. Development of micro-CT protocols for in vivo follow-up of mouse bone architecture without major radiation side effects. Bone, 2011; 49(4):613-622.

[26] Longo AB, Sacco SM, Salmon PL, Ward WE. Longitudinal use of micro-computed tomography does not alter microarchitecture of the proximal tibia in sham or ovariectomized sprague–dawley rats. Calcif Tissue Int, 2016; 98(6):631-641.

[27] Brouwers JEM, van Rietbergen B, Huiskes R. No effects of in vivo micro-CT radiation on structural parameters and bone marrow cells in proximal tibia of wistar rats detected after eight weekly scans. J Orthop Res, 2007; 25(10):1325-1332.

[28] Mustafy T, Benoit A, Londono I, Moldovan F, Villemure I. Can repeated in vivo micro-CT irradiation during adolescence alter bone microstructure, histomorphometry and longitudinal growth in a rodent model? PloS One, 2018; 13(11):e0207323-e0207323.

[29] Buie HR, Moore CP, Boyd SK. Postpubertal architectural developmental patterns differ between the L3 vertebra and proximal tibia in three inbred strains of mice. J Bone Miner Res, 2008; 23(12):2048-2059.

[30] Wehrle E, Tourolle né Betts DC, Kuhn GA, Scheuren AC, Hofmann S, Müller R. Evaluation of longitudinal time-lapsed in vivo micro-CT for monitoring fracture healing in mouse femur defect models. Sci Rep, 2019; 9(1):17445.

[31] Xu H, Lu BW, Zheng BJ, Tian J, Qi B, Deng YX, He ZZ, Su DS, Wang XR. Smaller sized inhaled anesthetics have more potency on Senescence-Accelerated Prone-8 mice compared with Senescence-Resistant-1 mice. J Alzheimers Dis, 2014; 39(1):29-34.

[32] Li XM, Su F, Ji MH, Zhang GF, Qiu LL, Jia M, Gao J, Xie ZC, Yang JJ. Disruption of hippocampal Neuregulin 1-ErbB4 signaling contributes to the hippocampus-dependent cognitive impairment induced by isoflurane in aged mice. Anesthesiology, 2014; 121(1):79-88.

[33] Lambers FM, Schulte FA, Kuhn G, Webster DJ, Müller R. Mouse tail vertebrae adapt to cyclic mechanical loading by increasing bone formation rate and decreasing bone resorption rate as shown by time-lapsed in vivo imaging of dynamic bone morphometry. Bone, 2011; 49(6):1340-1350.

[34] Scheuren AC, Hulst G, Kuhn GA, Masschelein E, Wehrle E, De Bock K, Müller R. Hallmarks of frailty and osteosarcopenia in prematurely aged PolgA$^{D257A/D257A}$ mice. bioRxiv, 2020:758243.

[35] Scheuren AC, Hulst G, Kuhn GA, Masschelein E, Wehrle E, De Bock K, Müller R. Hallmarks of frailty and osteosarcopenia in prematurely aged PolgA$^{D257A/D257A}$ mice. J Cachexia Sarcopenia Muscle, 2020:Forthcoming.

[36] Kujoth GC, Hiona A, Pugh TD, Someya S, Panzer K, Wohlgemuth SE, Hofer T, Seo AY, Sullivan R, Jobling WA, Morrow JD, Van Remmen H, Sedivy JM, Yamasoba T, Tanokura M, Weindruch R, Leeuwenburgh C, Prolla TA. Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. Science, 2005; 309(5733):481.

[37] Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, Bohlooly YM, Gidlof S, Oldfors A, Wibom R, Tornell J, Jacobs HT, Larsson NG. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature, 2004; 429(6990):417-423.

573 [38] Bellantuono I, de Cabo R, Ehninger D, Di Germanio C, Lawrie A, Miller J, Mitchell SJ,
574 Navas-Enamorado I, Potter PK, Tchkonia T, Trejo JL, Lamming DW. A toolbox for the
575 longitudinal assessment of healthspan in aging mice. Nat Protoc, 2020.
576 [39] Whitehead JC, Hildebrand BA, Sun M, Rockwood MR, Rose RA, Rockwood K,
577 Howlett SE. A clinical frailty index in aging mice: comparisons with frailty index data
578 in humans. J Gerontol A Biol Sci Med Sci, 2014; 69(6):621-632.
579 [40] Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in linear
580 mixed effects models. J Stat Softw, 2017; 82(13):1-26.
581 [41] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power
582 analysis program for the social, behavioral, and biomedical sciences. Behav Res
583 Methods, 2007; 39(2):175-191.
584 [42] Geurts J, Nasi S, Distel P, Müller-Gerbl M, Prolla TA, Kujoth GC, Walker UA, Hügle
585 T. Prematurely aging mitochondrial DNA mutator mice display subchondral osteopenia
586 and chondrocyte hypertrophy without further osteoarthritis features. Sci Rep, 2020;
587 10(1):1296.
588 [43] Eriksen EF. Normal and pathological remodeling of human trabecular bone: three
589 dimensional reconstruction of the remodeling sequence in normals and in metabolic
590 bone disease. Endocr Rev, 1986; 7(4):379-408.
591 [44] Robey PG,Bianco P, *CHAPTER 14 - Cellular mechanisms of age-related bone loss*, in
592 *The Aging Skeleton*, Clifford J. Rosen, Julie Glowacki, and John P. Bilezikian, Editors.
593 1999, Academic Press: San Diego. p. 145-157.
594 [45] Bouxsein ML, Myers KS, Shultz KL, Donahue LR, Rosen CJ, Beamer WG.
595 Ovariectomy-induced bone loss varies among inbred strains of mice. J Bone Miner Res,
596 2005; 20(7):1085-1092.

597 [46] Willey JS, Livingston EW, Robbins ME, Bourland JD, Tirado-Lee L, Smith-Sielicki H,
598 Bateman TA. Risedronate prevents early radiation-induced osteoporosis in mice at
599 multiple skeletal locations. Bone, 2010; 46(1):101-111.
600 [47] Bell EG, McAfee JG, Constable WC. Local radiation damage to bone and marrow
601 demonstrated by radioisotopic imaging. Radiology, 1969; 92(5):1083-1088.
602 [48] Pacheco R,Stock H. Effects of radiation on bone. Curr Osteoporos Rep, 2013;
603 11(4):299-304.

604

605

Supporting information

Table S1. The effects of imaging session number (group), genotype and the interaction effect (genotype*group) on bone morphometric parameters and frailty index (FI) at 40 weeks of age were compared via two-way ANOVA analysis. The p-values and effect sizes (*f*) of the main and interaction effects, respectively are listed below (significance level $\alpha = 0.05$).

Table S2. P-values, effect sizes (*f*) and achieved power obtained by cross-sectional (one-way ANOVA) and longitudinal analysis (paired t-test). (Significance level $\alpha = 0.05$).