Peer Review File

Manuscript Title: Aged Skeletal Stem Cells Generate an Inflammatory Degenerative Niche

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

Ambrosi, Marecic and McArdle et al. titled "Aged skeletal stem cells generate an inflammatory niche that impedes skeletal integrity" explores aging phenotype of the skeleton with a particular focus of the previously characterized (from the same group) skeletal stem cells (SSCs). The renal graft of isolated cells, and parabiosis experiments establish that the isolated cells from different time points exhibit different behavior that is likely refractory to environmental influence. Interestingly, the authors, based on the comparison between the isolated cells in the end suggest that it is still possible to change the aging phenotype by the simultaneous application of two factors, affecting purportedly two different facets of SSC activity, and restore certain aspects of skeletal regeneration in mice.

The aging phenomena of skeleton is an important subject for exploration and the authors utilize an array of experimental systems to tackle the subject. Characterisation of intrinsic aging of the isolated cells (Figure 2) and contrasting the systemic effects on aging of SSCs and HSCs by parabiosis system (Figure 3) give good reference points that will be of interest to many readers.

However, several points are not adequately addressed in the manuscript that raises concerns barring acceptance of the manuscript in the current form.

First, it is unclear how potential heterogeneity within the combinatorial cell surface marker defined SSC population will impact the interpretation of their data. Even comparing the cell surface marker definition of SSCs to the group's previous paper (Chan et al. 2015), CD200 has been dropped, which distinguishes SSCs from pre-BCSPs. Moreover, their own single cell analysis is presented to show that there is diversity in the SSC population that is "lost" in older population, emphasizing heterogeneity as an aspect of SSC in young bones. This begs the question whether the isolated SSCs consist rather of lineage-restricted subpopulation of cells and aging process entails reduction of particular subpopulation rather than intrinsic changes of individual SSC cells. Analyzing the subclusters (leiden cluster 2, 3, 5 shown in Extended Figure 5a) might be useful to address this aspect.

In addition, more robust characterization of the stemness of the isolated SSCs from adult by clonal culture as described in their previous report (Chan et al. 2015) might provide necessary controls that the cell surface marker definition is sufficient to determine the skeletal stem cell population. Importantly, the authors do not directly provide evidence that the aging phenotype of the skeleton is mediated by the aging SSCs. What they show is that the isolated aged SSCs exhibit different functional behavior in heterotopic condition. Thus it remains a possibility that there might be other cell populations that are not captured by the cell surface markers that could potentially exhibit stem cell behaviors and responsible of differences in the regenerative process. A negative control where SSCs and BCSPs are depleted to show no cells without the markers are capable of showing functional output would be a way to address this issue.

On a similar note, the application of BMP2 and aCsf1 could have directly influenced the osteoblasts and/or osteoclasts to generate phenotypic differences in bone regeneration (Figure 6c,d,e), in parallel, not as a consequence of the changes observed of SSC and BCSP output (Figure 6b). At least providing an assay of clonality measurement as done in Figure 1a would be helpful to more properly link the effect of

BMP2 and aCsf1 on SSCs and link to the regenerative phenotype.

Lastly, some critical methodological details are omitted. Most serious is the entire lack of processing information for single cell analysis (described more below). There is also omission of details in statistical treatment of microarray analysis, alongside with absence of critical information regarding test of equal variance.

Specific comments

(Line 29-30) It is rather tenuous to claim that there is a gradual outcompetition between young SSCs and old SSCs. The only suggestive information is the UMAP plot from the single cell analysis of three time points (P3, 2-month, 24-months). Outcompetition implies that different subpopulations of SSCs compete for same resources, which may or may not be true and need at least subclusering analysis of the leiden clusters the authors present.

Although overall, Figure 1 shows that both the regenerative phenotype (assessed by clonality, but see minor comments) as well as the surface marker defined stem cell or multipotent cell population number and cellular characteristics (proliferation, apoptosis, subset marked by CD49f as well as in vitro differentiation) are age-associated, whether they are "coupled" (line 124) is unclear.

The reviewer also notes that in ref. 2 from the same group, they described mSSC as CD51+Thy1-6C3-CD105-CD200+ cells, making a distinction of the presence of CD200 to the CD200-negative pre-BCSPs, which are not described as self-renewing stem cells (ref.2 Figure 2G). Given the FACS strategy employed in the manuscript, the omission of CD200 gating leaves the possibility that the described SSCs in these datasets are still heterogeneous.

Single-cell analysis

(Line 203-204) As for the single cell analysis, important details how the clustering and differential gene expression analysis is conducted to derive Figure 4a, b are missing. This is rather exemplified by the Figure panel 4a, where the legend says that "leiden clustering" has been conducted, but there is no actual clustering result available aside from the UMAP embedding and sample coloration, a rather clear oversight (the intended figure appears to be Extended Figure 5b). The reviewer notes that the graph-based clustering algorithm generates assigned cell clusters, which can also differ in terms of choice of resolution parameter for the clustering algorithm, as well as the preprocessing parameters such as how many principal components were used, what kind of potential batch effect removal has been conducted. Those details should be presented in the methods, in addition to robustness validation that the choice of such parameters are reasonable in concluding heterogeneity of SSCs.

Also the reviewer notes that after quality control procedure, 167 cells from 2-month SSCs, and only half for 24-month SSCs (88) remain, so the apparent "homogeneity" of 24-month SSCs may arise due to sample capture efficiency, or by systematic sequencing depth differences coupled with an arbitrary cut-off of 500 genes per cell threshold (Line 852). If possible, showing additional controls, for example that per-cell RNA content between the three groups do not differ significantly would help to alleviate such concerns.

If it is indeed the case that for 2-month SSCs there is heterogeneity that is lost in 24-month SSCs (given the rather lower gene count and small cell numbers, it would be rather difficult to convince readers without proper statistical analysis), the follow up analysis (Figure 4b, c, d, e, f) should all consider this heterogeneity.

For example, are the differentially expressed gene analysis driven by the fact that the heterogenous composition of 2-month SSCs, or does one also observe differences of gene expression between 2-month SSCs and 24-months SSCs that are clustered together? The Figure 4b and extended Figure 5 may seem to suggest the latter, but it is not entirely clear in the text. In addition, the proposition of diversity/heterogeneity found by single cell analysis will affect the interpretation of all the quantifications

put forward in Figure 1, 2, 3.

Figure 5a. The microarray analysis lacks statistical information of replicate number and thus confidence of the fold change comparisons. In the main text (Line 237) it is stated that Rankl is increased, at least based on the coloring scheme is is more like that Rankl is decreased in Thy1, 6C3 population, and entirely judged by the color, SSC/BCSP "increase" of Rankl seems to be not so prominent. There is actually no way to judge what difference may be significant.

Whereas the genetic model and statistics for Figure 5f is clearly convincing (but see Figure 7b comment below), it remains unclear whether PBS hydrogel would be a proper control (a proper control would be a recombinant protein similar to Csf1 without binding affinity to Csf1r) for Figure 5e experiments. Given the high variance of reading and borderline p-value (0.0455) for the read out for Extended data figure 7b suggests that the assay is underpowered to firmly establish that Csf1 exposure impairs recovery.

Combinatorial targeting of the aged SSC differentiation

The Figure 6 is an important experiment with Figure 6e showing the recovery by the combinatorial application of two factors in fracture, however, similar positive control threshold line should be presented also in Figure 6b and Extended Figure 8a-d. Without a benchmark of young bone repair, it is hard to assess how relevant the application of various factors are in relation to the age effect.

Discussion

(Line 307) The usage of words like "symbiotic" is irrelevant to the claims in the manuscript.

(Line 310) The authors have not shown that the same surface marker defined cells are SSCs, and their interpretation regarding the single-cell studies rather raises the concern that the surface markers may not be sufficient to define bona fide cell population per se.

(Line 315) The phrase that "SSC aging is multifactorial" and "modulation of hematopoietic crosstalk" is a bit convoluted for readers, given the frequent suggestions from the HSCs but the author's intention most likely being modulation of osteoclast activity. it might be straightforward to describe autonomous and non-autonomous activity of SSCs in maintaining the balance between osteogenic output and osteoclast activity.

(Line 320-322) Given suggestions of pro-inflammatory cytokines in the text (Line 32, 180-181), it might be of interest how these gene expression compares to the senescence associated secretory phenotype (SASP). The authors only mention once the term senescence in their discussion but proper references and discussion might be of interest to general audience.

Minor comments

(Line 80) The manuscript mentions plural reports, but only reference 15 is presented.

(Line 104-105) "Given earlier reports showing the SSC's role in skeletal maintenance and regeneration" citation is missing.

(Line 105) "SSC activity" is poorly defined in the text.

(Line 140-141) Any points to actual data figure for "formed larger ossicles" need to be added.

(Line 251) Principal component analysis cluster shows that one of the 24-month old SSCs are actually closer to the 2month group than the other two. To claim this clustered "distinctly" is rather questionable.

(Line 285) There is no clear reference to the aCsf1 inhibitory effect in the setting tested mainly in Figure 6. The two reference 2 and 37 do not describe it, so it is necessary to properly show that aCsf1 agent is

working as intended.

(Line 366-370) Discussion is entirely speculative with poor connection to the finding presented by the authors. It should be removed.

(Line 670) proper citation (ref 2.).

(Line 855-857) Although important details regarding the preprocessing of the single cell data is presented, the information on the actual analysis of generating the plots were put into a single sentence. Proper citation for the software (scanpy) should be put into the manuscript.

Figure 1d. Although the figure legend says that the distribution is derived from six different callus regions, the figure is presented in a way that does not properly provide the statistics. Neither in the figure legend nor the method section contains statistics such as the number of animals used, how many clones counted, most importantly, the figure does not show the variation of the individual samples in terms of clonality. A formal statistical test on distribution (such as chi-squared) may be performed.

Figure 1e and f. Although the figure show strong differences between the two age groups, there is a great imbalance of replicate numbers between 2-month old (11 femurs in total, probably from 6 animals?) and 24-month old group (3 femurs in total, probably from 2 animals? Unclear). Although Figure 1e and 1f show results from different experimental animals (uninjured vs fracture) they show same group numbers (11 vs 3) -- it is hard to imagine that such differences in number was planned ahead, what was the rationale for the imbalance?

Specifically, given the apparent differences in variance between the two groups (also present for Figure 1g), the usual assumption of homoscedasticity does not seem to be appropriate. Were the results statistically tested for equal variance?

Moreover, given both e and f display raw number of cells (only loosely normalized by a skeleton (femur) or size-unspecified callus), it naturally raises the concern of batch effect or confounding due to processing. It would be helpful if the authors could provide controls to ensure the dissociation condition is similar between the two time point seniors. For example, would the difference in cell counts still maintained when it is normalized to the total number of cells (CD45-TER119-ITAGV+Tie2-) isolated and run under FACS? At the least, to alleviate concerns the authors need to provide more complete information regarding the numbers in legend and/or method section (number of batches, animal counts, number of cells isolated etc.).

Figure 1k., the reviewer is curious what the denominator means. Are they total CD45-TER119- cells? (Same question applies for Figure 1h) It is difficult to find either in the legend or method section. Given the importance of this panel in supporting the claim that the aged and young SSCs have different lineage output, clarification is warranted. At face value, it appears that after six days of culture, a vast majority of cells after seeding are neither SSC, BCSP, Thy1+ or 6C3+ cells.

Figure 2c. The variance of 2-month aged animals and 24-month animals seem to be different. How as the unpaired t-test conducted? For default t-test functions, equal variation assumption will be clearly violated, as such, the calculation of p-value would be affected.

Figure 3c. Again, not completely clear what the denominator is for the frequency of alive (%) y-axis.

Figure 4c & Extended Figure 5a. It would be better to explain what the track plots is for readers not familiar with this type of plot. More better, it may be consolidated with Figure 4b by showing on the x axis the marker genes for Figure 4c.

Figure 5b. The figure legend is not entirely clear in what the connections between the ligand expressing cell subpopulation and the receptor expressing cell population entail. The unit for numbers in each rounded boxes are also not described in the figure legend, nor is it described in the method section

clearly.

Figure 6b. Rather than placing colors to the dots, it would be much better to put another x axis label with the designated experimental condition to aid the readers.

Figure 6d. The point shapes are all messed up, and the reviewer has no ability to distinguish which group corresponds to the top line (hexagon).

Extended Figure 2. CD51 == ITAGV but this is not clearly stated in the figure legend.

Extended Figure 3d. It is hard to tell any difference between 2-month panel and 24-month panel, although the text (Line 132-133) suggests that only 24-month SSCs do not undergo adipogenic differentiation.

Extended Figure 4e. The color scheme is reversed (red minus, blue plus), contra to Extended Figure 6a, Figure 5a color scheme. This has to be fixed.

Extended Figure 5b. As mentioned above, this seems to be the proper leiden clustering results for Figure 4a. It would be more informative for readers though to present the actual frequency (in terms of percentage or better, the actual numbers) of each age group for the clusters, preferably as stacked box plots or similar graphical representation rather than UMAP plot.

Extended Figure 6k. Inclusion of human specimen (Extended Figure 6k), is cursory, lacking functional characterization of the SSCs in adults. At least either provide similarity in transcriptome (since this is a microarray study and bulk, single cell RNA-seq as well as microarray datasets are presented in this manuscript) to suggest conserved signaling axis (Line 258-259) between mice and humans, or leave this out for future discussion.

Extended Figure 7d,e. It appears that the BV/TV is highly variable in Csf1-KO animals, and in wildtype the mechanical strength, BMD, Trab. Th. all have sexual dimorphism raising concerns whether the statistical comparison is properly powered. The reviewer is curious how the equal-variance assumption has been met to conduct unpaired student's t-tests to derive the p-values.

Referee #2 (Remarks to the Author):

The manuscript entitled "Aged skeletal stem cells generate an inflammatory niche…" describes an impressive set of experiments demonstrating the age-dependent properties of a population defined by the authors as skeletal stem cells. The paper explores these cells' molecular properties, regeneration potential and interaction with the hematopoietic system using a series of powerful and well-designed experiments. First, the authors demonstrate aged bones regeneration to be involving a smaller number of distinct clones (and overall much smaller SC output) than observed in younger mice. Then, the authors beautifully dissect intrinsic and systematic properties of this age-related effect, using in-vitro assay, renal grafts and parabiosis. The conclusions from this part of the paper are that aged bones are intrinsically impaired in their regeneration potential, possibly due to some lower potential for stem cell expansion. This part of the paper is excellent in my mind, but there are some questions regarding the linkage between clonaity and ageing phenotype that must be better explained and discussed as detailed below.
The authors then try to characterize the molecular properties of the aged-SSC population using scRNA-seq and some additional bulk-RNA-seq, giving rise to some candidate differential markers among which is Csf1. Csf1 function in SSC and bone-regeneration is the focus of the reminder of the paper. The authors convincingly demonstrate Csf1 is overexpressed in aged healing bones, and that Csf1 is affecting negatively healing in-vitro. They also show mice that are haplodeficient for Csf1 had stronger femora which is consistent with a negative role for high Csf1 in maintaining bone structure. But following fracture, Csf1+/- bones regenerate poorly, indicating possible complex feedback and dose-dependent effects are modulating these phenotypes. Finally, the authors demonstrate that that bmp2 treatment of

aged mice fractures is improved by adding low Csf1 dose. The second part of the paper uses a powerful array of techniques, and end up with encouraging demonstration of potential treatment, but the basis for some of the arguments is weaker. In particular, it will be important to provide better delineation of the different possible sub-population of Csf1 expressing or non-expressing SSCs – both in young and aged mice as suggested below. It is also important to get more clarity and robustness regarding the effect of adding low Csf1 to the already effective Bmp treatment, and ideally examine the effect this may have on young bones.
In conclusion, however, I think this is an excellent paper, well written and showing intriguing new results on a very timely and important question.

Major comments:
1. Clonality and ageing in the bone. The data in figure 1 is very interesting, but it is difficult to link the age related clonality to the other observations in the paper. The authors should provide some clarity about 2 conflicting hypotheses: a) that aged bones are just losing stem cells with age, resulting in less clones, less regeneration, but per single SSC a similar behavior (e.g. as may be suggested by figure 1 but not the rest of the paper) b) that the pool of bone SSC is ageing "synchronously" such that collectively their potential for regeneration is lower (as suggested by Figure 4-5). Do the authors suggest that both the total number of SSC is going down with age, and their regeneration potential as a group is also decreasing? Is there quantitative evidence that can prove this? (ie. The size of a typical clone in the Rainbow assay is stable with age?)
2 In age related clonal hematopoiesis, it is frequently suggested that some recurrent mutation are driving the take-over of clones over HSC niches. It should be clarified if the authors wish to argue a similar process is occurring in the bone or else the belief is that senescence/epigenetic deterioration is the driving force. This is assuming the authors have excluded the possibility of homeostasis with the aged hematopoietic system as a major driver. I find it difficult to understand a prolonged process of ageing as involving homogeneous transition from state A (young) to state B (old). A much more intuitive model is of a mixture of states (A and B) and shifted balance between the two, as sometime argued to occur in the HSC/MPP/GMP/CMP/CLP differentiation hierarchy.
3. The analysis of gene expression analysis is somewhat shallow, and its display and explanation in the Methods section is lacking. Figure 4a shows heterogeneity between cells, which is not investigated by the authors. It is only briefly explained by saying that the aged SSC become more homogeneous, which is a very partial explanation. Judging from the figure, it is possible that there is a shift between SSC states during aging, which may provide more insight into the functional decline described in the paper. The only analysis performed on the single-cell data is for differential gene expression, which could have also been performed on bulk data.
It is suggested that the authors will re-analyze their scRNA-seq data to try and characterize variation within and between group, and in particular support their Csf1 analysis downstream. At least the following should be done and reported:
- show QC metric for the single cells – percent mitochondrial, ribosomal and stress (heat shock, proteasome) transcripts per cell, grouped by age. For example, the authors interpret the genes c-Jun, Jund and Junb as associated with commitment to osteochondrogenic fates. These genes are involved in many stress-related processes, often associated also with Irf7/Interferon response and more. One must make sure the young scRNA-seq data is not simply more affected by stress. In that respect it is essential to show batch information for scRNA-seq and make sure it is controlled for.
- show cell cycle analysis (expression of S-phase and M-phase genes) – to demonstrate proliferation rates of the single cell population.
- present a heatmap of de-novo single cell clusters and show the distribution of expression of the genes best separating the clusters. Label cells in each cluster by their age. Perhaps this can lead the authors to some new insight into the regulation of e.g. Csf1 in young mice, or to the existence of "aged" like SSC in younger time points.
Naturally the above analysis would benefit from generating additional scRNA-seq data. But I do not suggest this should be mandatory. Proper analysis can already improve the results.
4. It is not clear how the expression values in Figures 4e-f (and similarly in Extended Data Figure 5c-d) are calculated, and they appear to exaggerate differential expression. For example, Figure 4b shows a very subtle difference in Csf1 expression, while 4f shows a very distinct one. The authors should instead show boxplots or dotplots for key markers, with careful statistics and normalization to library depth and

QC as discussed above.
One possible approach for more sensitive analysis is to search for SSC genes that are correlated with Csf1 levels within each age group. It is not a-priori clear that top correlated genes would be those that are also differentially expressed in young/aged mice.
5. The Methods section should provide more detail on how the data were analyzed, and the code for the analysis should be made publicly available (before publication).

6. The authors link the bias of aged SSC to Thy1+ stroma to the shift of HSCs to the myeloid lineage in aging (e.g. "our findings indicate that the aged SSC lineage favors the selective expansion of myeloid-skewed HSCs"). These conclusions are not adequately supported in the paper. Even if previous work showed that Thy1+ supports myeloid cells, and this work shows that SSCs create more Thy1+ cells in aging, it is not shown that this increase leads to more myeloid cells in aging. The only evidence presented here for this conclusion is the upregulation of csf1, the concomitant csf1r by myeloids, and the GO enrichment of "macrophage migration", and this evidence is largely anecdotal. Other evidence for the effect of SSC on hematopoiesis, such as the high expression of Cxcl12, Kitl and Vcam1 in aging SCCs is also not convincing. These results are interesting and can be briefly mentioned, but no strong conclusions can be drawn from them.
7. The combinatorial treatment data is really interesting, but the high variance in the overall effect (Fig 6e) makes it difficult to argue convincingly that Bmp2 treatment is improving when adding aCsf1^low. Statistics must be improved here. One can suggest (but I don't suggest this should be required) to perform some additional follow ups (for example gene expression of Bmp2 treated SSCs with and without Csf1low and Csf1high). The authors are heavily relying on the results to motivate their story – it is important to make sure the result is robust.
Typos / small errors:
- According to the legend of Figure 4a, the figure shows Leiden clustering of the cells, but the cells are marked by the mouse age.
- "SSCs could also contributes" – should be "contribute" (page 15).

Referee #3 (Remarks to the Author):

In this manuscript Ambrosi et al. evaluate whether aging associated decreases in bone mass are due to SSC intrinsic pathology and evaluate the mechanisms involved. The topic addressed is both novel and important, as while there have been extensive studies of mechanisms involved with aging-associated skeletal pathologies, none of this prior work have resolved the basis of aging associated skeletal pathologies down to discrete mesenchymal populations. However, relatively little data is offered that directly maps aging associated cellular dysfunction to defects in SSCs. A similar concern is that the primary mechanism identified, increased CSF1 secretion, is a well known regulator of osteoclastogenesis and that the experiments performed cannot distinguish a role of CSF1 in SSC related mechanisms of skeletal aging versus the general and known pro-osteoclastogenic functions of CSF1. While the methods used here are technically sound, there are several major concerns that the approaches applied do not appear appropriate to address the questions posed or are unable to adequately support the conclusions proposed. Specific details are described below.

Major points

1. Prior reports on human and murine SSCs identify these cells as growth plate resident cells. However, here the properties of these cells are being studied under aging conditions where the growth plate is "lost", especially in humans where there is essentially complete growth plate fusion. This apparent conflict requires clarification.

2. While the data on transplantation of SSCs and BCSPs post fracture to evaluate the size and composition of the in vivo ossicles they form in a transplantation system is important, it cannot be excluded that this result reflects an attenuation of the ability of non-stem progenitors generated from aged SSCs to progress to mature cell fates. The finding that BCSPs also display a similar attenuation of

ossicle formation capacity and in vitro differentiation capacity raises the possibility that some of the effects seen occur in a "post-SSC" cellular compartment or reflect a broad and general dysfunction of mesenchymal cells not specific for SSCs. The data currently offered does not clearly deconvolute the relative contribution of these SSC versus "post-SSC" mechanisms. An important aspect of performing this deconvolution is to address the degree to which aging specifically influences SSC stemness properties, such as by assessing relative self renewal capacity of aged versus young SSCs in a competitive or limiting dilution system. Along similar lines, the only experiment directly addressing whether aging results in differences in the cellular output specifically of SSCs is conducted in vitro in Extended Fig 2e, which only shows modest differences. In vitro differentiation is well known to not necessarily reflect in vivo differentiation capacity in bone. It will be important to specifically assess the cellular output of aged vs young SSCs versus BCSPs and pre-BCSPs in vivo to address the stem cell specific nature of aging induced dysfunction. This data is needed to justify the claims made in the section heading (line 78-9).

3. A distinct but related issue is that the SSCs examined at older ages may represent a qualitatively different cell type (as opposed to the interpretation favored in the manuscript that aged SSCs represent the same cells as young SSCs, albeit with quantitative alterations in mature tissue production capacity). This may simply reflect that the flow panel applied here is not able to resolve the compositional transitions occurring with aging within the SSC compartment. This concern is supported by the scRNA-seq data in Fig 4a, b and Extended data fig 5a, b showing qualitative changes in makeup of the SSC pool, with multiple clusters present at young ages and only a singer cluster present in older mice. The omission of CD200 from the analysis (see major point 4 below) also raises the level of concern that unappreciated heterogeneity may be central to the observations made here. Indeed, it would seem likely the changes in cell types comprising the SSC pool as suggested by these scRNA seq studies would be central to the aging phenotypes observed. However, these changes in the cell types falling within the SSC pool are not clearly described and their potential significance to the aging associated phenotypes is not examined experimentally. Specifically, finding that Cxcl12 is increased in 24mo SSCs suggests that this pool may be now be comprised of Cxcl12 abundant reticular (CAR)/LEPR+ cells that are normally not present within the Chan et al. murine SSC definition, as CAR cells are marrow stromal elements separate from cell populations in the growth plate region.

4. In comparison with the prior Chan et al. report on murine SSCs, CD200 has been excluded from the definition used here. This omission is important as it represents a substantial change in the SSC definition. SSCs here are only designated as Lin-CD105-ITAGV+ cells, a much larger and presumably more heterogenous population than the CD200+ subset of the above. This scheme also omits the previously described “pre-BCSP” category and merges these cells with SSCs. Are CD200+ SSCs present in aged mice or does the pool of cells designated as SSCs here predominately reflect pre-BCSPs in aged mice? This issue has important implications as to whether the defects observed here truly reflect SSC dysfunction.

5. Dynamic histomorphometry is needed for the experiments in Fig 3b, c to assess the degree to which these results here reflect differences in anabolic versus catabolic capacity. Supplementing this with serum levels of CTX or a similar catabolic marker is helpful to provide an assessment of catabolic activity. This question of anabolic versus catabolic nature of the differences observed is not addressed by the bone marrow transplant study performed in Fig 3j, as differences in catabolic activity can also be intrinsic to alterations in the osteoclastogenic capacity of the mesenchymal compartment or the influence of other non-circulating cell types present in bone on osteoclastogenesis.

6. In many of the mouse studies, the "n"s employed seem fairly small given the expected variation in the models employed, such as in Fig 3 at 3-4 per group for many groups. How many independent repeats were performed and how many mice were analyzed across all repeats in total? Review of the total data would be helpful in establishing confidence in these results. Similar comments apply to Fig 5e. It is unclear as presented whether the increase in BMD comparing HA to IA samples in Fig 3b is biologically meaningful or statistically significant, and the statistics for this comparison should be clarified. The p values provided appear to be comparing the HA and IA samples to the IY samples, but this is likely not the most relevant comparison for these groups. Instead the comparison of greatest interest is comparing

HA to IA specimens.

7. Bone mass phenotypes are expressed in many figures as BMD instead of the more standard trabecular BV/TV and cortical thickness. Does the data in Fig 3b and 3j represent uCT or use of a DXA-like platform? Use of uCT is important given that DXA measurements are relatively insensitive and may mask important phenotypes. Cortical and trabecular phenotypes should be separately reported using standard nomenclature for uCT, as there are compartment specific differences in response, and phenotypes in cortical thickness have a more direct relevance to human fracture risk. At least for a subset of key experiments, the fracture phenotypes throughout would be better reported as an analysis of total callus bone volume as determined by uCT as opposed to the 2D radiograph based callus index.

8. While there is some nuance to the data, looking at the results in Xiong et al. Nature Communications 2018 in mice with sheddase-resistant RANKL in comparison with mice with complete RANKL or RANK deficiency overall emphasizes that the role of soluble RANKL in bone metabolism is fairly minor. Thus, the lack of differences in soluble RANKL are of unclear relevance to the phenotypes observed and don't necessarily offer evidence that non-RANKL factors are involved. Measurement of tissue-level RANKL expression would be more relevant in this context.

9. The interpretation offered for Fig 3h, that myeloid skewing of HSCs reflects altered stromal signaling/that myeloid skewing with aging is an SSC property that can be rescued in HA parabionts doesn't appear to be supported by the current experiments. It is unclear how it can be excluded that the phenotype seen in the HA vs IA parabionts reflect direct actions of young parabiont derived circulating factors on HSCs as opposed to the stroma. The system used with transplantation of these HSCs to young recipients would appear to favor the former, as these effects persist post-HSC transplantation. It may be more informative to examine the HA versus HY derived HSCs directly in the context of the bones of either the HA versus HY parabionts to determine to what degree any myeloid skewing phenotypes reflect cell intrinsic effects of aging on the local environment versus effects mediated by circulating factors.

10. The experiment supplementing bone marrow cultures with RANKL and CSF1 (extended data fig 6, lines 261-5) does not show that increases in Csf1 expression and secretion are an important factor contributing to the increased osteoclastogenic capacity of aged mesenchymal cells. Instead, this data shows an increased osteoclastogenic potential of the aged bone marrow stroma, which may be due to any number of factors, including differences in osteoclast precursors, differences in the receptiveness of these precursors to CSF1/RANKL or differences in the amount of co-stimulatory factors present that enhance osteoclast differentiation. Additionally, no direct evidence is offered to functionally establish that any aging-associated osteoclastogenic phenotypes are specifically attributable to SSC dysfunction as opposed to dysfunction in more mature osteoblasts, and the overall contribution of SSCs as sources of osteoclastogenic signals is unclear.

11. Similarly, it is well established that Csf1 is a key osteoclastogenic factor and that the outcomes seen in Fig 5e-h all reflect the known and “generic” roles of Csf1 in osteoclastogenesis in these settings. These experiments do not specifically implicate CSF1 as a mediator of aging associated SSC dysfunction. Similar concerns apply to Fig 6a, as these results reflect the known general effects of BMP2 and CSF1 blockade and not a specific interaction indicating aging associated stem cell dysfunction.

12. The claim in the title is not clearly supported by the data presented. It is clear that aged skeletal stem cells display dysfunction in terms of osteogenic output, however, no direct experimental evidence is offered showing that an inflammatory niche is causal for these defects. Similarly no functional data is presented showing that SSCs are themselves the source of inflammatory factors crucial to aging associated skeletal pathologies.

Minor points:

1. Page 5, line 99, the language used implies that the prior published papers on human and mice identify human and murine versions of the same cell type. However, given the substantial differences in surface

immunophenotype between these cell types and also possible functional differences, it is unclear if these two reports are converge on a single cell type with species specific differences or distinct cell types in each study. Commentary and an update of the language used here would be critical in avoiding confusion about the current state of evidence regarding these two important studies. This issue may also have implications for the human relevance of this study that will be important to acknowledge.

2. The studies in the rainbow mice are helpful supporting data, but clonal clusters as suggested by this method only indicate clonal proliferation, which is not necessarily a stem cell property. This result could reflect proliferation of any later non-stem progenitor in the differentiation hierarchy. Updating the language on page 6 line 107 ("stem cell activity") is needed to reflect this.

3. Clarification of whether a true loss of the growth plate "page 5, line 82-3" versus an attenuation is observed at the ages examined would be helpful for placing results in context. Typically murine systems are considered as displaying a progressive attenuation but not a full loss of the growth plate with aging, such as in Mizuhashi et al. Nature 2018.

4. It is difficult to interpret extended fig 2d given that the high numbers in the Thy1+ fracture samples compress the axis for all other samples. Separate graphs or a broken axis are needed.

5. Why were NSG hosts used for study of murine SSCs? Typically that profound degree of immunodeficiency isn't necessary for this study design. This is a relevant issue as inflammatory mechanisms of SSC dysfunction are highlighted as a mechanism and may be impacted in NSG hosts.

6. The emphasis in interpretation in the Fig 3 data is that young blood does not rescue aged SSC cellular defects. However, the most striking effect seen overall is that, conversely, old blood appears to be sufficient to confer cellular dysfunction into young SSCs (such as comparing IY to HY in Fig 3b-d). This undercuts the major conclusion of this section of the manuscript, that "Instead, cell intrinsic changes to SSC activity seem to be the primary driver in the age-related functional decline of the skeleton", as cell extrinsic factors indeed appear to be sufficient to drive SSC dysfunction.

7. The emphasis in the text describing Fig 3 on whether or not HA parabionts are fully restored to the point of being similar to IY controls is puzzling, as HA parabionts could display substantial rescue without being comparable to IY controls. Comparison to IA controls would seem more relevant.

8. It is unclear how the conclusion that "SSC aging might therefore be the consequence of a priming event that institutes an intrinsically aged state which renders these cells henceforth insensitive to blood-borne factors," follows from the immediately preceding statement that circulating factors exert a negative effect on the SSC lineage. Conversely, it would seem that the most straightforward interpretation of the Fig 3 data would be that circulating cell extrinsic factors associated with aging are indeed sufficient to induce SSC dysfunction, but that circulating factors from young mice are not sufficient to reverse dysfunction in aged SSCs. The claim that aging is intrinsic to SSCs appears to be at odds with the SSC dysfunction observed in HY mice relative to IY controls in Fig 3b-d. This would imply a mixed model with both cell intrinsic and cell extrinsic components to the SSC aging phenotype. Another possibility is that while there are circulating factors that confer aspects of "oldness" to SSCs, there may simply not be corresponding circulating factors that can confer "youngness".

9. Regarding the description of cluster associated genes in page 10 line 209-11, Jund is a suppressor of osteogenesis as in Kawamata et al. J Cell Biochem, which would appear to run counter to the description in the text.

10. BMP2 plays complex roles in both anabolic and catabolic responses, so simply describing it as a gene associated with bone formation as on page 11 line 217 is potentially misleading. Similarly, the anabolic versus catabolic effects of TGF-beta on bone metabolism depend on amount and context (such as in Erlebacher et al, J Cell Biol 1996) in contrast with the statement on page 12 line 244.

11. Relevant to the statement that “Rankl by bone marrow stromal cells leads to enhanced

osteoclastogenesis and adipogenesis at the expense of osteogenesis" (p16, line 344-5), while RANKL levels may increase with aging and relative amounts of adipogenesis increase with aging, this reviewer is not aware of any clear or direct evidence that increases RANKL expression specifically drive the increases in adipogenesis at the expense of osteogenesis observed with aging.

12. Extended Data 2d and 2e are not mentioned in the main text.

13. In Figure 1d, how was the calculation for the number of clones made? The standard deviation is per group is missing in the plot.

14. In the main text Line 123, the figure is mislabeled as 1d-e, instead it should be 2d-e.

15. The TRAP staining image in Fig 2d is unclear. A better image, including a lower power view is needed.

Author Rebuttals to Initial Comments:

N/A

Reviewer Reports on the First Revision:

N/A

Author Rebuttals to First Revision:

Referee #1 (Remarks to the Author):

Ambrosi, Marecic and McArdle et al. titled “Aged skeletal stem cells generate an inflammatory niche that impedes skeletal integrity” explores aging phenotype of the skeleton with a particular focus of the previously characterized (from the same group) skeletal stem cells (SSCs). The renal graft of isolated cells, and parabiosis experiments establish that the isolated cells from different time points exhibit different behavior that is likely refractory to environmental influence. Interestingly, the authors, based on the comparison between the isolated cells in the end suggest that it is still possible to change the aging phenotype by thesimultaneous application of two factors, affecting purportedly two different facets of SSC activity, and restore certain aspects of skeletal regeneration in mice.

1. The aging phenomena of skeleton is an important subject for exploration and the authors utilize an array of experimental systems to tackle the subject. Characterisation of intrinsic aging of the isolated cells (Figure 2) and contrasting the systemic effects on aging of SSCs and HSCs by parabiosis system (Figure 3) give good reference points that will be of interest to many readers.

Response to comment 1: We would like to thank Reviewer 1 for the very positive and supportive remarks. As highlighted by the reviewer, our findings showed that aged SSCs acquire an intrinsically altered state impairing their normal function which cannot be rescued by concentrations in which pro-osteogenic factors are present in a youthful blood circulation. Despite that observation, and as pointed out as well by this reviewer, in the end we describeda combinatorial treatment that is able to reinstate SSC function in aged mice. We believe our ability to reinstate youthful SSC function is due to the fact that we locally apply factors (by

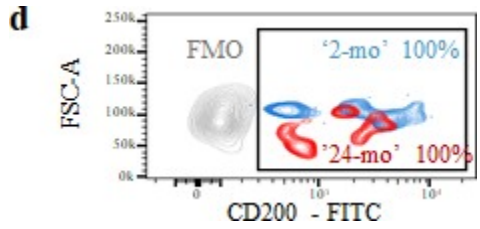
hydrogel) that alter the immediate niche environment and expose cells to a much more concentrated rejuvenation cocktail than old blood diluted with young blood can achieve.

However, several points are not adequately addressed in the manuscript that raises concerns barring acceptance of the manuscript in the current form.

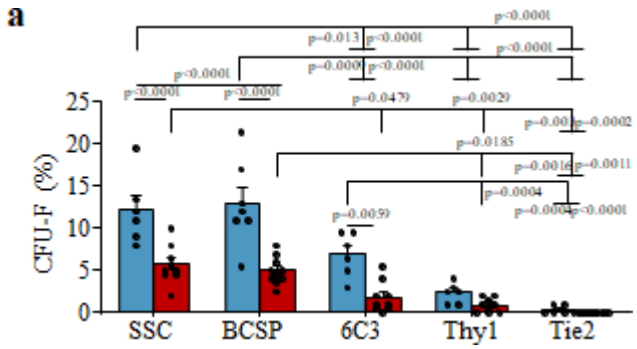
2. First, it is unclear how potential heterogeneity within the combinatorial cell surface marker defined SSC population will impact the interpretation of their data. Even comparing the cell surface marker definition of SSCs to the group's previous paper (Chan et al. 2015), CD200 has been dropped, which distinguishes SSCs from pre-BCSPs.

Response to comment 2: We thank the reviewer for the thoughtful comment. Our initial identification of the mouse skeletal stem cell indeed utilized CD200 as an additional criteria to distinguish SSC and pre-BCSP populations. However, this work was based on early development/fetal tissues (Chan et al. 2015, PMID: 25594184). In adult SSC lineages all CD51+Thy1-6C3-CD105- cells were found to be CD200 positive, thus not allowing further separation by this marker. We now include FACS analysis comparing young adult (2-mo) and

aged (24-mo) mouse bone single cell solution gated for the CD51+Thy1-6C3-CD105- cell population showing that virtually all adult SSCs express CD200 (lines 107-110; Extended Data Fig. 2d):



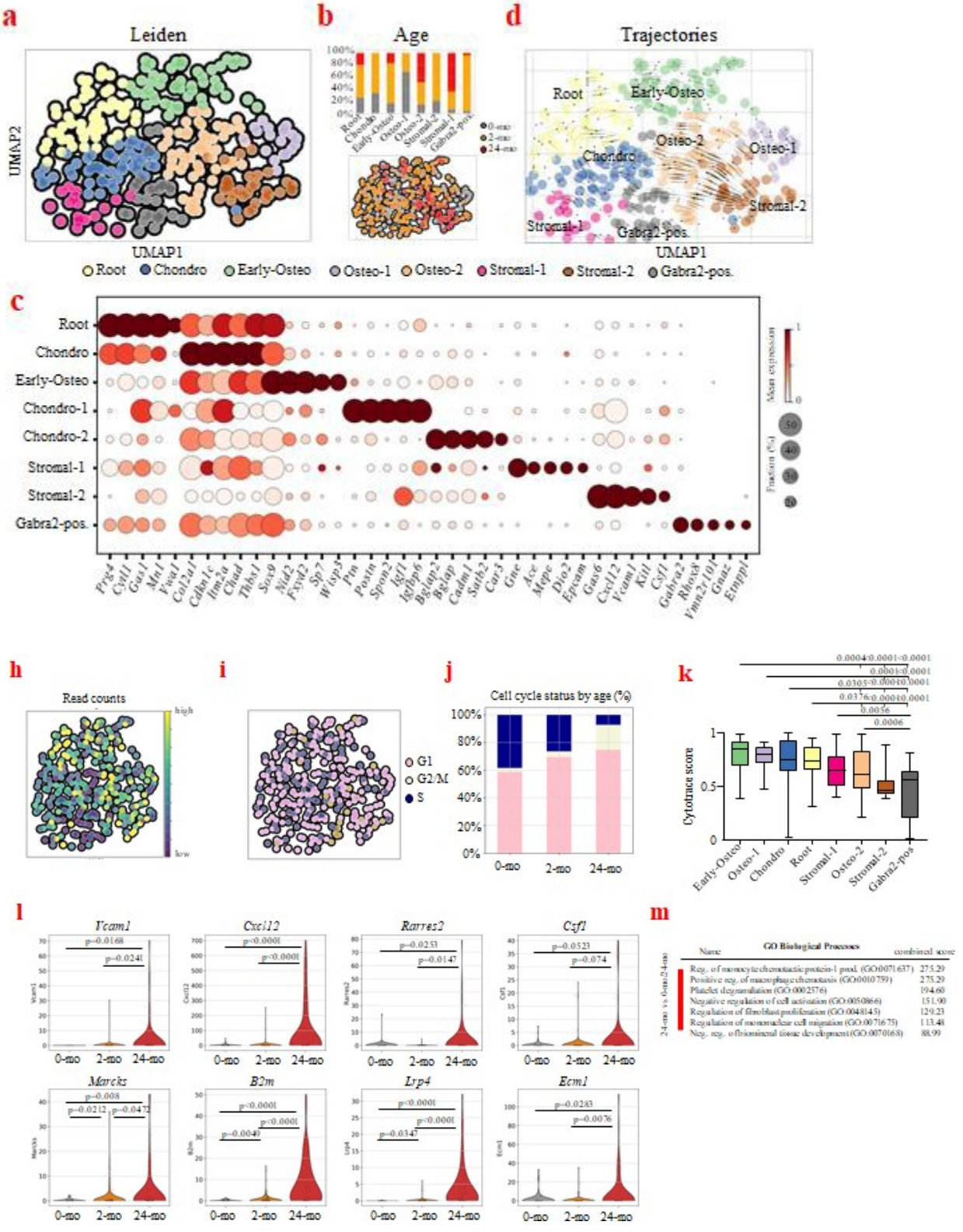
We have additionally conducted CFU-F assays on SSCs and previously defined downstream populations to show that phenotypic SSCs and BCSPs maintain the highest clonogenic potential throughout age, i.e. stem cell characteristics (Extended Data Fig. 3a).



Our previously presented data also showed that aged SSCs remain multipotent with age in vitro and in vivo, albeit at a lower extent (Fig. 1l-n). It is also apparent that the reduced colony forming potential could be passed down to downstream lineage restricted cell populations of the SSC, thus contributing to the age-related impairments.

3. Moreover, their own single cell analysis is presented to show that there is diversity in the SSC population that is "lost" in older population, emphasizing heterogeneity as an aspect of SSC in young bones. This begs the question whether the isolated SSCs consist rather of lineage-restricted subpopulation of cells and aging process entails reduction of particular subpopulation rather than intrinsic changes of individual SSC cells. Analyzing the subclusters (leiden cluster 2, 3, 5 shown in Extended Figure 5a) might be useful to address this aspect.

Response to comment 3: Per the reviewer's suggestions we have also re-analyzed the single cell RNA-Seq data to more closely examine SSC heterogeneity. Altogether SSCs of all age-groups show striking transcriptomic similarity but subtle differences reveal that 24-moSSCs display potentially altered pre-primed trajectories with pro-hematopoietic/inflammatory and anti-chondrogenic lineage directions. In young SSCs on the other hand gene expressionreveals higher skeletogenic potential as well as responsiveness to external stress-inducing factors (lines 245-285; Fig. 3a-d and Extended Data Fig. 5h-m).

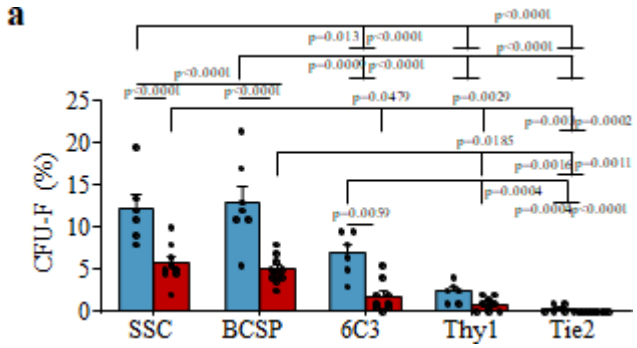


The clustering results suggest that SSCs might be made up of populations with varying developmental potential, similar to the hematopoietic system where HSCs can be divided into short and long term HSC as well as give rise to multipotent progenitor populations (MPPs)

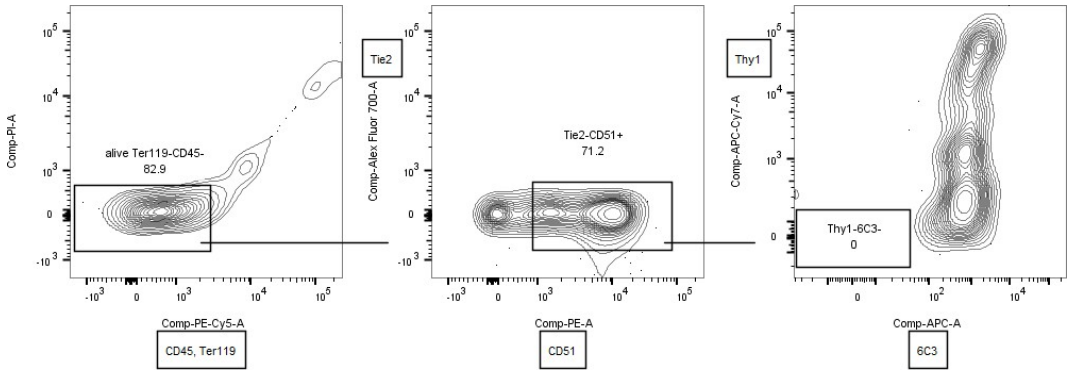
that show differences in lineage output (PMID: 31201709). Future studies are needed to further characterize these subpopulations which should lead to a more refined SSC lineage tree. Nevertheless, our combined results underscore the specificity of the used markers for the SSC population, which is an advance over plastic adhesion based methods and single gene markers used to purify 'MSCs'.

4. In addition, more robust characterization of the stemness of the isolated SSCs from adult by clonal culture as described in their previous report (Chan et al. 2015) might provide necessary controls that the cell surface marker definition is sufficient to determine the skeletal stem cell population. Importantly, the authors do not directly provide evidence that the aging phenotype of the skeleton is mediated by the aging SSCs. What they show is that the isolated aged SSCs exhibit different functional behavior in heterotopic condition. Thus it remains a possibility that there might be other cell populations that are not captured by the cell surface markers that could potentially exhibit stem cell behaviors and responsible of differences in the regenerative process. A negative control where SSCs and BCSPs are depleted to show no cells without the markers are capable of showing functional output would be a way to address this issue.

Response to comment 4: Since SSCs and BCSPs give rise to downstream progenitor populations of bone cartilage, and stromal cell types of the skeleton, the loss of functional activity at the apex of the lineage tree (SSCs/BCSPs) is very likely the major contributor to age-related skeletal impairments. Phenotypic SSCs/BCSPs remain the population with the highest clonogenic potential throughout mouse life and in aged bones (Extended Data Fig. 3a).



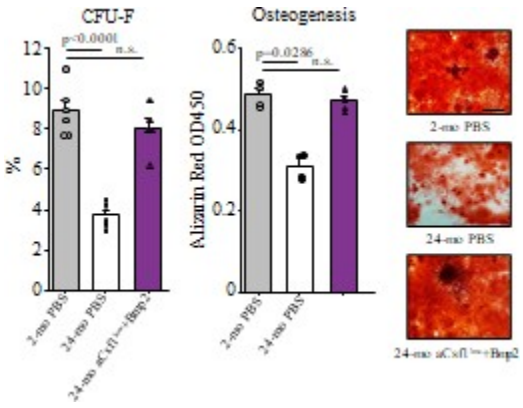
This is consistent with our previous findings showing that purified cell populations other than SSCs and BCSPs do not contribute to multi-lineage output in the transplant setting (Chan etal 2013 and 2015, PMID: 26216955 and 25594184), while at the same time cell properties and lineage output are altered at the stem cell level in aged mice (Fig. 1k-n, Extended Data Fig. 2e-f). Of note, SSC/BCSP depleted cells do not give rise to SSC/BCSPs upon in 7-daysin vitro culture (data not included in manuscript):



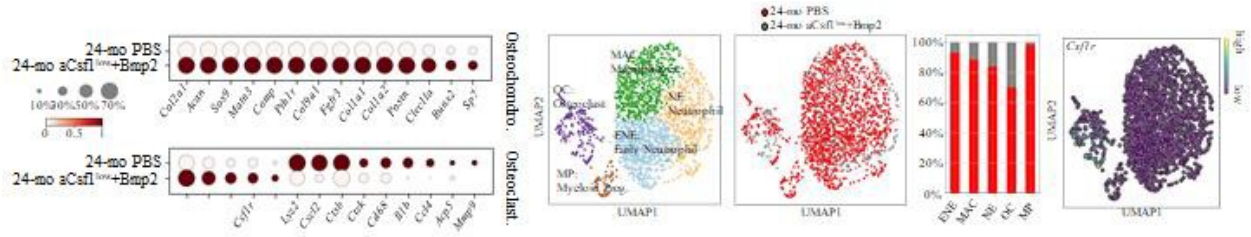
Altogether, our data strongly suggests that the major age-related bone-forming deficiencies originate at the SSC level, but this does not exclude the possibility that downstreampopulations derived from SSCs also inherit age-related stem cell defects. As our data showsSSC-derived stromal cells likely also contribute to skeletal dysfunctions in aged bones (Fig. 3e). Future work is needed to determine if downstream cell types inherit functional genetic and epigenetic changes from aged SSCs that contribute to the aging phenotype.

5. On a similar note, the application of BMP2 and aCsf1 could have directly influenced the osteoblasts and/or osteoclasts to generate phenotypic differences in bone regeneration (Figure 6c,d,e), in parallel, not as a consequence of the changes observed of SSC and BCSP output (Figure 6b). At least providing an assay of clonality measurement as done in Figure 1a would be helpful to more properly link the effect of BMP2 and aCsf1 on SSCs and link to the regenerative phenotype.

Response to comment 5: We agree with the reviewer that Bmp2 and aCsf1 could also have affected osteoblast and/or osteoclast to generate the phenotypic differences seen. We have noted this possibility in the discussion: “Closer examination of molecular changes in these fracture callus tissues support the notion that other cell types aside from SSCs, e.g., osteoblasts and immune cells, are positively affected by the rescue treatment.” (lines 435-437). However, we also note that aging corresponds to significantly diminished injury-activated expansion of SSC. When we previously suppressed injury-activated SSC expansion by irradiation (Chan et al 2015, PMID: 25594184), we observed that fracture repair was significantly impaired as in aged mice. We have now added additional n’s for our rescue experiments in Figure 4 (new version of manuscript) showing that BMP2 rescues age-diminished injury-activated expansion of SSCs. We have also performed additional clonality measurement as requested. These results are presented in Fig. 4d,f-g and show a strong connection between factor-rescued SSC frequency and functional CFU-F and osteogenesis readouts. In fact, combinatorial treatment of aged fracture with low aCsf1 and rBmp2 to propelled SSCs restore functionality to youthful (2mo) levels.



We also have conducted 10X single cell RNAseq analysis on day-10 fracture callus specimens of aged PBS-treated or aCSFlow/rBMP2-treated groups to explore molecular differences (Fig 4h-m and Extended Data Fig. 9). The results show that combinatorial treatment of fractures enhances osteochondrogenic gene expression in the callus and reduces osteoclast differentiation/number in support of the other results.



6. Lastly, some critical methodological details are omitted. Most serious is the entire lack of processing information for single cell analysis (described more below). There is also omission of details in statistical treatment of microarray analysis, alongside with absence of critical information regarding test of equal variance.

Response to comment 6: We have expanded the method section to include more detailed information. The full processing information is now provided on single cell RNAseq processing (lines 991-1066) and statistical analysis (lines 1066-1075). For microarray datasets each cell population was derived from bones of a pool of 3-5 mice. The described and validated Gene Expression Commons Database GEXC (PMID: 22815738) was used to analyze microarray samples. In brief, raw data was normalized and computed against the Common Reference, a large collection (n = 11,939) of publicly available mouse microarray data from the National Center for Biotechnology Information Gene Expression Omnibus. Meta-analysis of the Common Reference also provides the dynamic range of each probe set on the array. Only gene probesets with dynamic ranges >6 were selected and in situations where multiple probesets for a gene exist, the probeset with the widest dynamic range was used. While the results of the microarray analysis not only clearly support other transcriptomic and functional data, the fidelity of these microarrays has been validated by multiple studies (e.g., PMID: 30518860, 30150396, 26863982, 30241615, 25594184, 32537886).

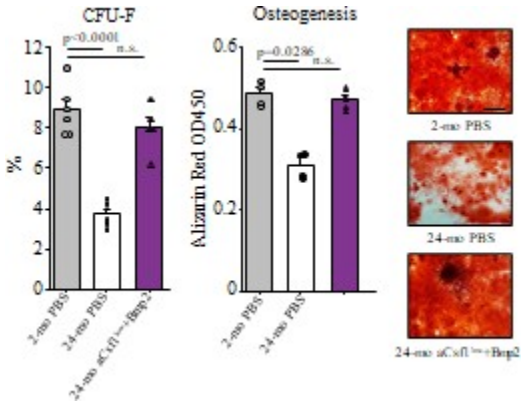
Specific comments

7. (Line 29-30) It is rather tenuous to claim that there is a gradual outcompetition between young SSCs and old SSCs. The only suggestive information is the UMAP plot from the single cell analysis of three time points (P3, 2-month, 24-months). Outcompetition implies that different subpopulations of SSCs compete for same resources, which may or may not be true and need at least subclusering analysis of the leiden clusters the authors present.

Response to comment 7: We acknowledge the reviewer's comment and have rephrased our conclusion to more appropriately adhere to what the presented data reflects. It now reads: "Single-cell transcriptomic studies tie the functional loss to a diminished transcriptomicdiversity of SSCs in aged mice corresponding to bone marrow niche transformation." (lines 30-32). Future studies will have to investigate the exact mechanism, e.g., competitionbetween SSC variants by establishing new types of SSC-variant specific reporter strains.

8. Although overall, Figure 1 shows that both the regenerative phenotype (assessed by clonality, but see minor comments) as well as the surface marker defined stem cell or multipotent cell population number and cellular characteristics (proliferation, apoptosis, subset marked by CD49f as well as in vitro differentiation) are age-associated, whether they are "coupled" (line 124) is unclear.

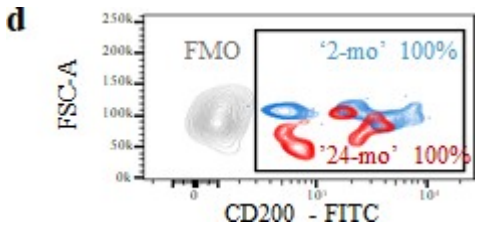
Response to comment 8: We also here have rephrased to "corresponds to" (line 131). In any case, it is reasonable to conclude that the decline in bone mass and the frequency of stem and progenitor cells are connected as shown throughout the whole manuscript, including the rescue of SSC activation and regeneration from fracture injury in factors-treatedaged mice (Fig. 4f-g).



9. The reviewer also notes that in ref. 2 from the same group, they described mSSC as CD51+Thy1-6C3-CD105-CD200+ cells, making a distinction of the presence of CD200 to the CD200-negative pre-BCSPs, which are not described as self-renewing stem cells (ref.2 Figure 2G). Given the FACS strategy employed in the manuscript, the omission of CD200

gating leaves the possibility that the described SSCs in these datasets are still heterogeneous.

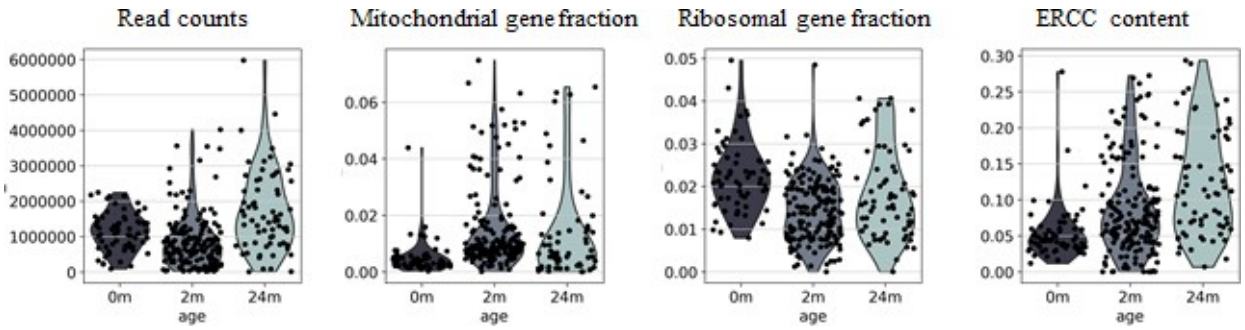
Response to comment 9: Please see comment 1 above and our response which addresses this point. CD200 does not separate SSCs in adult skeletal tissue (Extended Data Fig. 2d).



Single-cell analysis

10. (Line 203-204) As for the single cell analysis, important details how the clustering and differential gene expression analysis is conducted to derive Figure 4a, b are missing. This is rather exemplified by the Figure panel 4a, where the legend says that “leiden clustering” has been conducted, but there is no actual clustering result available aside from the UMAP embedding and sample coloration, a rather clear oversight (the intended figure appears to be Extended Figure 5b). The reviewer notes that the graph-based clustering algorithm generates assigned cell clusters, which can also differ in terms of choice of resolution parameter for the clustering algorithm, as well as the preprocessing parameters such as how many principal components were used, what kind of potential batch effect removal has been conducted. Those details should be presented in the methods, in addition to robustness validation that the choice of such parameters are reasonable in concluding heterogeneity of SSCs.

Response to comment 10: We thank the reviewer for this comment. We have included detailed methods on processing and selection of scRNAseq analysis parameters in the updated methods section (lines lines 991-1066). We also provide all information to reproduce the analysis as supplemental material (raw data has been GEO deposited and processing notebook has been made available as described in the Method section). We have used stringent quality control (QC) parameters to only include high quality single cell data with low mitochondrial and ribosomal gene content as well as low RNA-spike in mix control abundance:

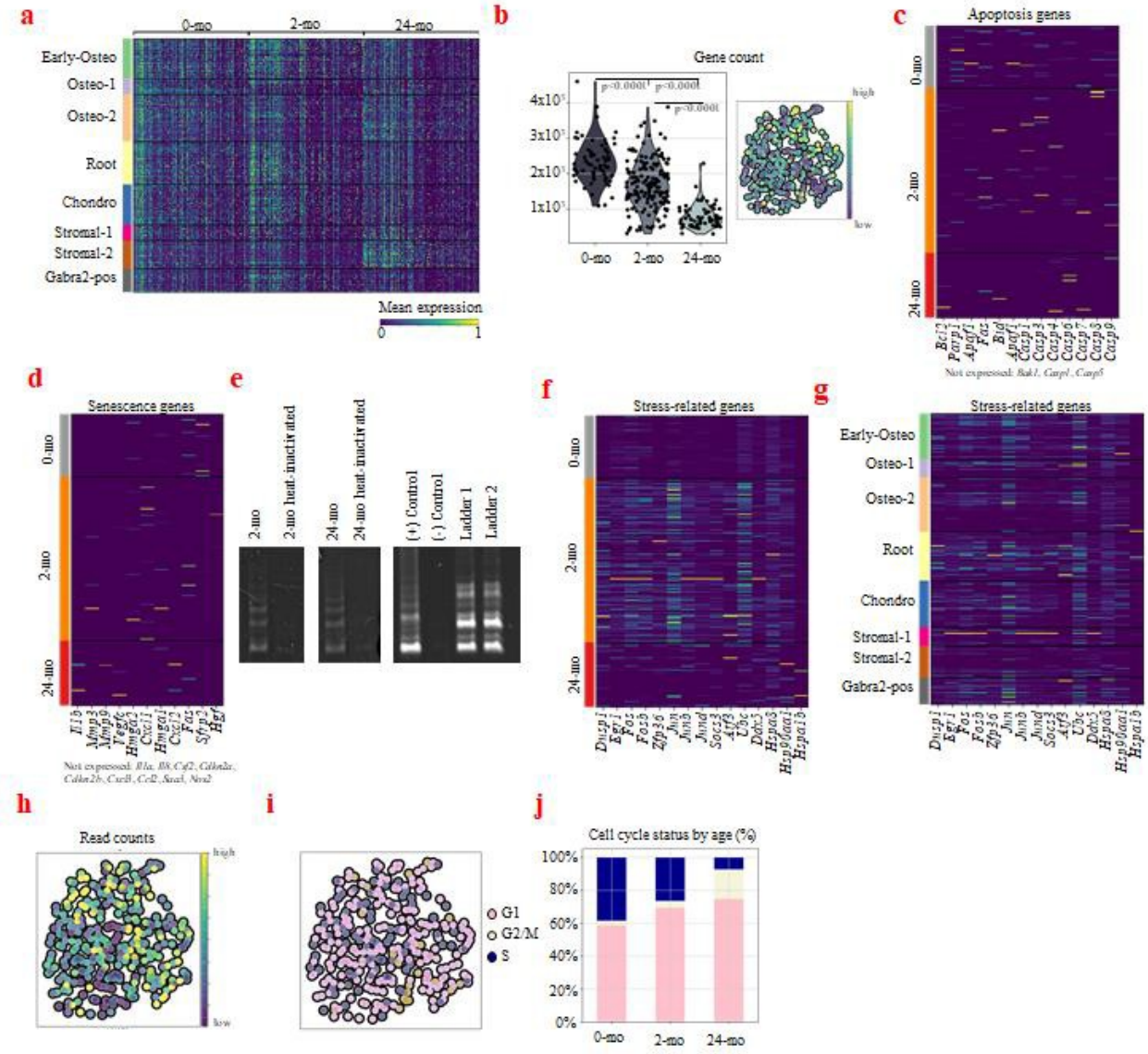


Additionally we have performed batch correction (Combat) for differences between processed single cell plates. We found that transcriptomic differences between single SSCs from 0-mo, 2-mo, and 24-mo mice are subtle. Varying the number of principal components (PCs) for our clustering analysis did not significantly change cell separation. Based on the PC elbow plot we settled on 36 PCs by which we were able to identify specific differences between SSCs of different age groups (Fig. 3a-d & Ext. Data Fig. 5, see also response to comment 1).

11. Also the reviewer notes that after quality control procedure, 167 cells from 2-month SSCs, and only half for 24-month SSCs (88) remain, so the apparent “homogeneity” of 24-month SSCs may arise due to sample capture efficiency, or by systematic sequencing depth differences coupled with an arbitrary cut-off of 500 genes per cell threshold (Line 852). If possible, showing additional controls, for example that per-cell RNA content between the three groups do not differ significantly would help to alleviate such concerns.

Response to comment 11: We have updated our single cell dataset and included details on QC (lines 1025-1035 and see comment 8 above) and results of quality assessment between different age groups (also see processing notebook linked in methods section). The differences in cells that passed very strict QC (see methods) in young versus aged SSCs aremultifaceted. First, the number of wells sorted per age group and capture rates varied by experiment. For example, capture rate for aged SSCs was lower as observed by lower numbers of wells with cDNA traces analyzed upon single-cell processing (data not shown). This was likely a technical issue at our facility since viability of SSCs during sorting did not affect the majority of cells during experiments (Apoptosis assay: >97% of cells are non- apoptotic Fig.1j.; no wide expression of apoptotic markers in single cell data, see Extended Data Fig. 5c; re-analysis of double-sorted SSC, not shown). Furthermore, scRNAseq analysisshowed that young cells are more receptive to processing-induced expression of stress- related genes compared to newborn and aged SSCs (Extended Data Fig. 5f-g). This might also indicate why young SSC lineage cells were more responsive to aged blood and 24-mo SSCs remained largely unaffected. Importantly, high expression of stress-related genes did not affect Leiden clustering, similar to cell cycle status. Aged SSCs did not express high levelsof genes related to senescence (or displayed changes in telomerase activity) further suggesting epigenetic changes underlying differences in their biology (Extended Data Fig. 5d-e, i-j). After conducting batch correction and looking at cell cycle status, counts per cell

(sequencing depth), and controlling for low quality cells as well as RNA-content we found that neither of these factors affected cluster analysis in the remaining cells. Thus, the fact that aged cells lose overall gene expression might be reflecting their overall greater epigenetic age and that the loss of stem cell potential corresponded to reduced overall gene expressionperhaps as a result of epigenetic silencing, see Gulati et al. and their Cytotrace algorithm aswell as Zhang et al. (PMID: 31974247, PMID: 32020082). Based on these observations andadditional CytoTrace analysis we concluded that significantly lower total gene counts in 24- mo SSCs are likely biological and reflective of reduced developmental potential.

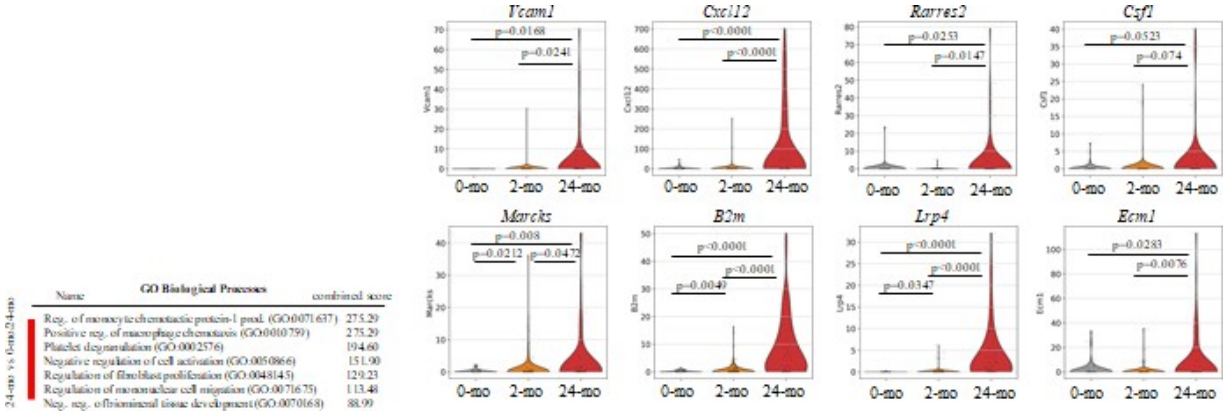


12. If it is indeed the case that for 2-month SSCs there is heterogeneity that is lost in 24-month SSCs (given the rather lower gene count and small cell numbers, it would be rather difficult to convince readers without proper statistical analysis), the follow up analysis (Figure 4b, c, d, e, f) should all consider this heterogeneity.

Response to comment 12: Although we do not think that the variation in analyzed cell numbers contributed to the decreased diversity, we provide updated analysis that implies biological reasons for the loss of diversity. As elaborated above young SSCs seem to display higher developmental potential than aged SSCs and contain subclusters indicating enhanced potential to commit to skeletal lineage trajectories (Fig. 3a-d and Extended Data Fig. 5).

13. For example, are the differentially expressed gene analysis driven by the fact that the heterogenous composition of 2-month SSCs, or does one also observe differences of gene expression between 2-month SSCs and 24-months SSCs that are clustered together? The Figure 4b and extended Figure 5 may seem to suggest the latter, but it is not entirely clear in the text. In addition, the proposition of diversity/heterogeneity found by single cell analysis will affect the interpretation of all the quantifications put forward in Figure 1, 2, 3.

Response to comment 13: We have re-analyzed our dataset as mentioned before and also applied batch correction (Combat) to account for differences in processing/age groups. Our updated and more detailed Leiden cluster analysis now shows a rather homogeneous clustering of the three age groups, but also distinct clusters enriched for 24-mo SSCs. Specific differentially expressed genes between the different clusters as well as their age-composition are now shown in (Fig. 3 and Extended Data Figure 5) and gene lists for Leiden clusters and age groups are included as supplemental material (Supplemental Information 1). We have then extended the analysis to examine specific age-related differences and found enriched pro-hematopoietic, pro-monocytic gene expression patterns in 24-mo SSCs compared to younger counterparts (Extended Data Fig. 5l-m).

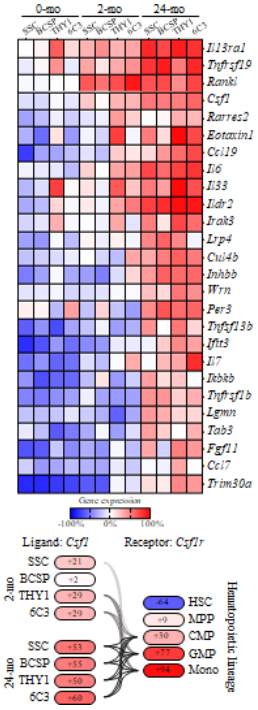


Overall, the discovered changes in diversity by scRNAseq correspond to differences in functional readouts between young and aged SSCs as shown in previous figures. Aged SSCs are skewed to commit towards a non-osteochondrogenic, stromal fate corresponding to their diminished ability to support bone formation and repair while potentially exacerbating higher bone resorption by skewing HSC and downstream lineages.

14. Figure 5a. The microarray analysis lacks statistical information of replicate number and thus confidence of the fold change comparisons. In the main text (Line 237) it is stated that Rankl is increased, at least based on the coloring scheme is is more like that Rankl is

decreased in Thy1, 6C3 population, and entirely judged by the color, SSC/BCSP “increase” of Rankl seems to be not so prominent. There is actually no way to judge what difference may be significant.

Response to comment 14: As mentioned above in response to comment 4, the microarray analysis was based on GEXC (PMID: 22815738) which provides expression values based on a dynamic range calculated from close to 12,000 publicly available reference mouse microarray datasets and which has been validated by many studies. We have changed presentation of values and included absolute values, rather than percentage change, to allow better interpretation of the data (Fig. 3e,k). *Rankl* gene expression remains similar in young and aged SSCs. The fact that *Rankl* and *Csf1* are the two crucial factors for osteoclastogenesis, and that *Csf1* specifically presents with the tendency to be upregulated in the aged SSC lineage was one of the main reasons we focused on Csf1 in the following analyses.



15. Whereas the genetic model and statistics for Figure 5f is clearly convincing (but see Figure 7b comment below), it remains unclear whether PBS hydrogel would be a proper control (a proper control would be a recombinant protein similar to Csf1 without binding affinity to Csf1r) for Figure 5e experiments. Given the high variance of reading and borderline p-value (0.0455) for the read out for Extended data figure 7b suggests that the assay is underpowered to firmly establish that Csf1 exposure impairs recovery.

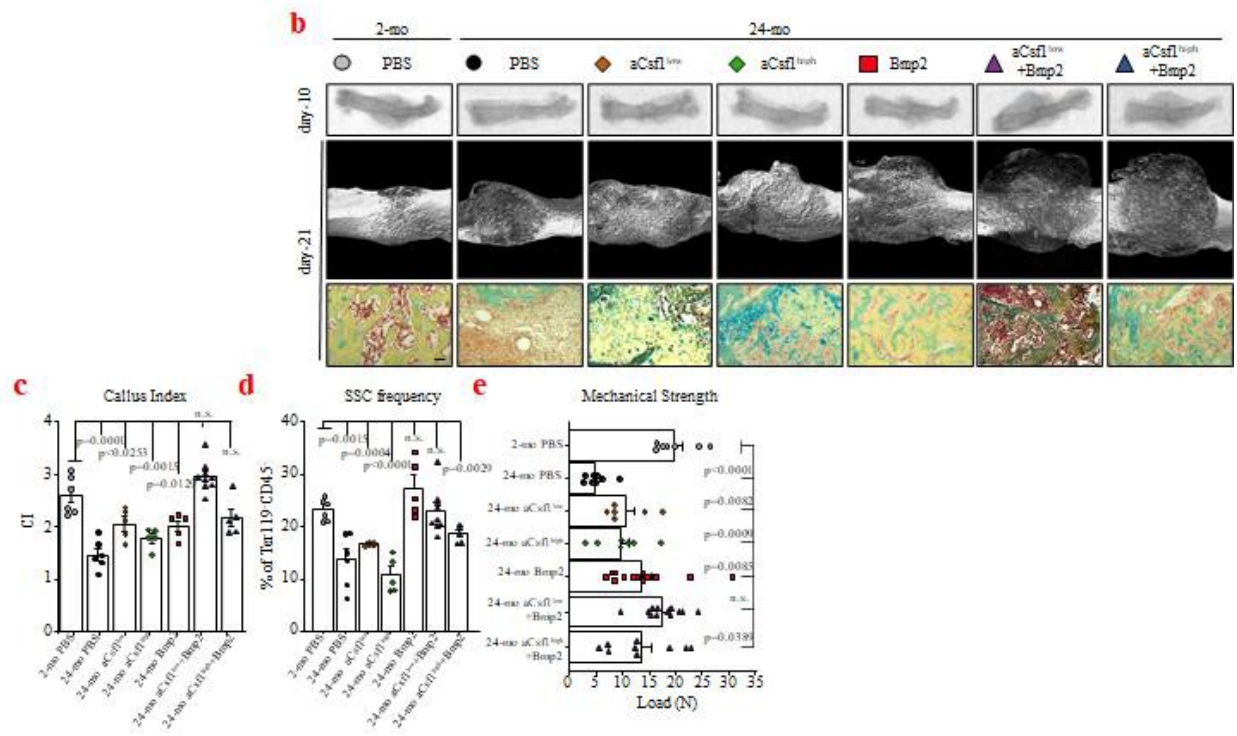
Response to comment 15: Fracture model experiments, even with stabilizing pins show biological variability, due to the inherent difficulty of the procedure and the effects of mouse

movement likely leading to higher variations within a group than other regeneration models. Fig. 5e (now Fig. 3o and Extended Data Fig. 7b-c) clearly shows one specific outlier for each treatment group driving the variation. Rather than excluding outliers, this variability has been accounted for in the statistical testing yielding a significant p-value. It would require a great many more aged mice for slight improvement in the p-value and we are limited in our accessto this precious resource and the conclusion are supported by additional experiments as presented in Figure 4.

16. ## Combinatorial targeting of the aged SSC differentiation

The Figure 6 is an important experiment with Figure 6e showing the recovery by the combinatorial application of two factors in fracture, however, similar positive control threshold line should be presented also in Figure 6b and Extended Figure 8a-d. Without a benchmark of young bone repair, it is hard to assess how relevant the application of various factors are in relation to the age effect.

Response to comment 16: We agree with the reviewer. We now have added additional animals for the rescue experiments, including a young PBS-treated reference control to show that the combinatorial treatment with low aCsf1 and rBmp2 specifically performs most closelyto what is seen in young fracture healing (Fig. 4b-e).



17. ## Discussion

(Line 307) The usage of words like “symbiotic” is irrelevant to the claims in the manuscript.

Response to comment 17: The phrasing has been altered to: “concerted” (line 375).

18. (Line 310) The authors have not shown that the same surface marker defined cells are SSCs, and their interpretation regarding the single-cell studies rather raises the concern that the surface markers may not be sufficient to define bona fide cell population per se.

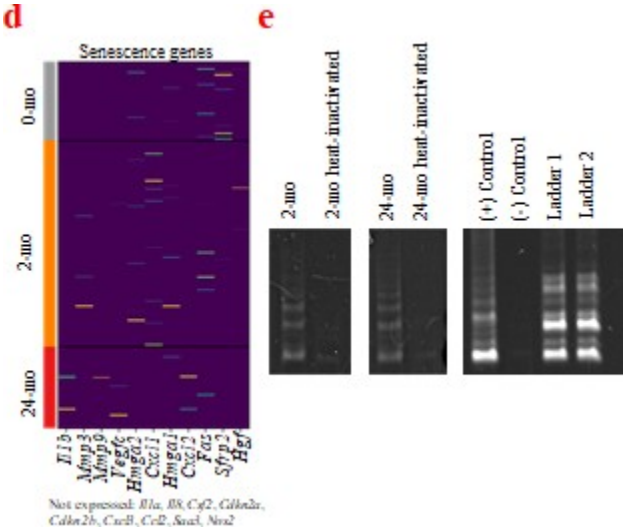
Response to comment 18: As described above provided functional and single cell RNA-sequencing data are supporting the notion that phenotypic SSCs of young and aged mice are targeted by the same surface marker expression (Fig. 3a-d & Extended Data Fig. 2d, 3a, 5a).

19. (Line 315) The phrase that “SSC aging is multifactorial” and “modulation of hematopoietic crosstalk” is a bit convoluted for readers, given the frequent suggestions from the HSCs but the author’s intention most likely being modulation of osteoclast activity. it might be straightforward to describe autonomous and non-autonomous activity of SSCs in maintaining the balance between osteogenic output and osteoclast activity.

Response to comment 19: We have now specified the description of our discoveries to more closely reflect the suggestions of the reviewer. “Lastly, targeting SSC autonomous and non-autonomous activities can restore aging-impaired regeneration by simultaneously enhancing local SSC lineage activity while finely modulating of pro-osteoclast activity.” (lines 381-383).

20. (Line 320-322) Given suggestions of pro-inflammatory cytokines in the text (Line 32, 180-181), it might be of interest how these gene expression compares to the senescence associated secretory phenotype (SASP). The authors only mention once the term senescence in their discussion but proper references and discussion might be of interest to general audience.

Response to comment 20: We have analyzed our single cell dataset for SASP related genes and did not find a clear signature of senescence-related genes in aged SSCs (Extended Data Fig. 5d). While some of the pro-hematopoietic markers, such as Cxcl12 and Csf1, sometimes are included in SASP profiles the lack of expression of other SASP specific genes argues for non-senescent related modes of actions. The lack of changes in telomerase activity between 2-mo and 24-mo SSCs further supported that (Extended Data Fig. 5e). Due to the ambiguityof the word “senescent” we have now changed it to “functionally distinct” (line 390).



Minor comments

Minor 1. (Line 80) The manuscript mentions plural reports, but only reference 15 is presented.

Response to minor comment 1: An additional study has been added as reference (Line 83).

Minor 2. (Line 104-105) “Given earlier reports showing the SSC’s role in skeletal maintenance and regeneration” citation is missing.

Response to minor comment 2: Citation has been added (line 110).

Minor 3. (Line 105) “SSC activity” is poorly defined in the text.

Response to minor comment 3: In line 99-101 it now states: “To determine if reduced stem cell activity, including proliferation and differentiation capacity, could be responsible for the aged skeletal phenotype in mice, we next focused on the cellular origins of bone-forming cell types.”

Minor 4. (Line 140-141) Any points to actual data figure for “formed larger ossicles” need to be added.

Response to minor comment 4: This was intended to refer to the same figures as cited in the preceding sentence (now Figure 1m and Extended Data Fig. 3e-f).

Minor 5. (Line 251) Principal component analysis cluster shows that one of the 24-month old SSCs are actually closer to the 2month group than the other two. To claim this clustered “distinctly” is rather questionable.

Response to minor comment 5: To avoid further controversy we have removed the principal component analysis from the revised version of this manuscript as it does not substantially contribute to the conclusions of the manuscript.

Minor 6. (Line 285) There is no clear reference to the aCsf1 inhibitory effect in the setting tested mainly in Figure 6. The two reference 2 and 37 do not describe it, so it is necessary toproperly show that aCsf1 agent is working as intended.

Response to minor comment 6: The antibody against Csf1/Mcsf (FisherScientific, Cat#MAB4161SP) has been validated by the manufacturer for its neutralizing actions against Csf1/Mcsf. Additionally, experimental results clearly show the anticipated outcome of Csf1 blockade, i.e. higher bone content/mass (Fig. 3 and 4).

Minor 7. (Line 366-370) Discussion is entirely speculative with poor connection to the finding presented by the authors. It should be removed.

Response to minor comment 7: We have removed this section.

Minor 8. (Line 670) proper citation (ref 2.).

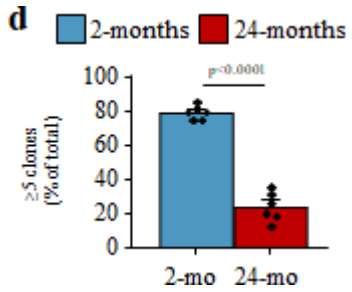
Response to minor comment 8: Citation has been corrected.

Minor 9. (Line 855-857) Although important details regarding the preprocessing of the single cell data is presented, the information on the actual analysis of generating the plots were put into a single sentence. Proper citation for the software (scanpy) should be put into the manuscript.

Response to minor comment 9: As mentioned above methods have been amended. The detailed paragraph on processing single cell RNA-sequencing data is now available in lines 991-1066. The processing notebook is also available and will eventually be added to the GEO accession (see method section).

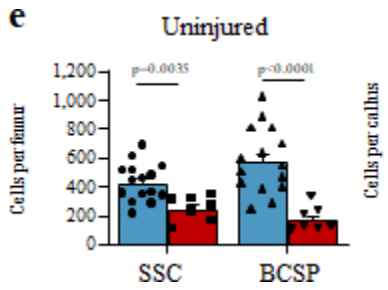
Minor 10. Figure 1d. Although the figure legend says that the distribution is derived from six different callus regions, the figure is presented in a way that does not properly provide the statistics. Neither in the figure legend nor the method section contains statistics such as the number of animals used, how many clones counted, most importantly, the figure does not show the variation of the individual samples in terms of clonality. A formal statistical test on distribution (such as chi-squared) may be performed.

Response to minor comment 10: We re-analyzed the data for this figure and now provide differences in the number of clones containing 5 or more cells of the same color that were counted for young and aged mice. The figure legend has been amended to state: “Six distinct callus regions (5-19 clones per section) from two mice per age group were counted.” (Fig. 1d)



Minor 11. Figure 1e and f. Although the figure show strong differences between the two age groups, there is a great imbalance of replicate numbers between 2-month old (11 femurs in total, probably from 6 animals?) and 24-month old group (3 femurs in total, probably from 2 animals? Unclear). Although Figure 1e and 1f show results from different experimental animals (uninjured vs fracture) they show same group numbers (11 vs 3) -- it is hard to imagine that such differences in number was planned ahead, what was the rationale for the imbalance?

Response to minor comment 11: Since sourcing of aged animals is quite difficult and costly, experiments were performed as aged mice became available. In some cases one experiment with a single aged mouse was accompanied with multiple younger counterparts for comparisons. Multiple experiments were combined for statistical comparisons. Each data point contains measurements of one single mouse. We have added additional n in this revision for Fig. 1e further supporting the previous observations.

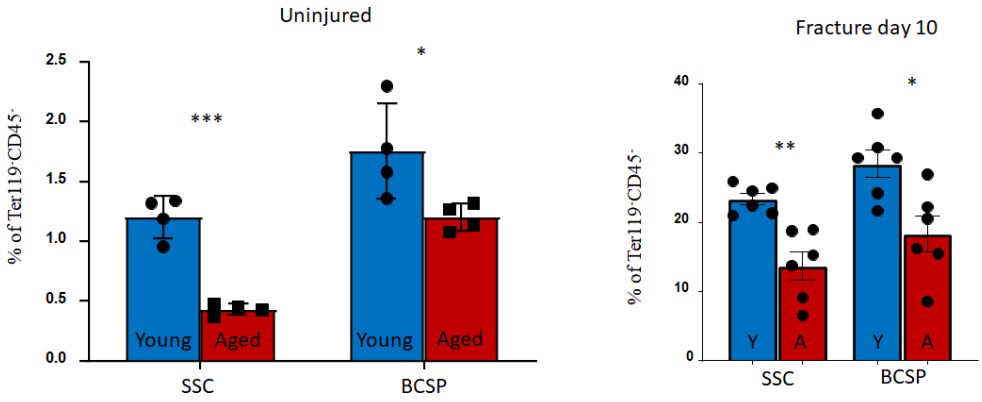


Minor 12. Specifically, given the apparent differences in variance between the two groups (also present for Figure 1g), the usual assumption of homoscedasticity does not seem to be appropriate. Were the results statistically tested for equal variance?

Response to minor comment 12: Statistical testing has been updated and now includes adjustment of t-test for non-normality (Mann-Whitney test) and unequal variances (Welch's correction) testing. See also updated method section lines 1068-1077.

Minor 13. Moreover, given both e and f display raw number of cells (only loosely normalized by a skeleton (femur) or size-unspecified callus), it naturally raises the concern of batch effector confounding due to processing. It would be helpful if the authors could provide controls to ensure the dissociation condition is similar between the two time point seniors. For example, would the difference in cell counts still maintained when it is normalized to the total number of cells (CD45-TER119-ITAGV+Tie2-) isolated and run under FACS? At the least, to alleviate concerns the authors need to provide more complete information regarding the numbers in legend and/or method section (number of batches, animal counts, number of cells isolated etc.).

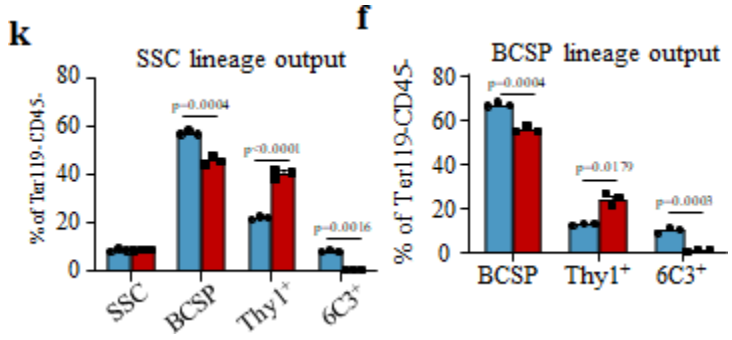
Response to minor comment 13: We have also analyzed frequencies of cells per CD45-TER119- showing the same trends but decided to display total counts per femur in these graphs as we believe this metric more accurately reflects differences in SSC abundance between different age groups. All presented data represent biological replicates of at least two independent experiments unless otherwise noted. Data presented here show that the same results are obtained when analyzing SSC/BCSP frequencies per Ter119-CD45- fraction.



Minor 14. Figure 1k., the reviewer is curious what the denominator means. Are they total CD45-TER119- cells? (Same question applies for Figure 1h) It is difficult to find either in thelegend or method section. Given the importance of this panel in supporting the claim that theaged and young SSCs have different lineage output, clarification is warranted. At face value,it appears that after six days of culture, a vast majority of cells after seeding are neither SSC,BCSP, Thy1+ or 6C3+ cells.

Response to minor comment 14: Results for these graphs are now shown as % of Ter119-CD45- cells and labeled on the y-axis. It becomes clear that standard culture media supplemented with FBS does not maintain the majority of seeded cells in a stem cell state

but rather induces spontaneous differentiation towards downstream lineages. Future studies will be needed to optimize media conditions for long-term expansion of phenotypic SSCs without committing to downstream fates as has recently been shown for HSCs (PMID: 31142833). Given that the in vitro conditions for young and aged SSCs are the same in these experiments, the results clearly show differences in the intrinsic properties between young and aged SSCs incl. differences in their lineage potential (Fig. 1k and Extended Data Fig. 2f).



Minor 15. Figure 2c. The variance of 2-month aged animals and 24-month animals seem to be different. How as the unpaired t-test conducted? For default t-test functions, equal variation assumption will be clearly violated, as such, the calculation of p-value would be affected.

Response to minor comment 15: This data has been reanalyzed with the appropriate Welch's correction of the t-test (Fig. 1n).

Minor 16. Figure 3c. Again, not completely clear what the denominator is for the frequency of alive (%) y-axis.

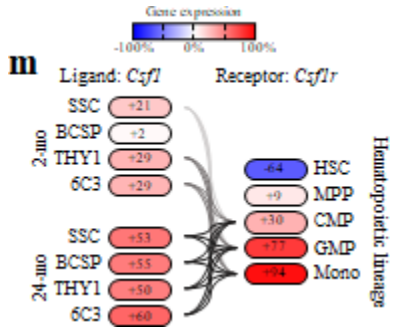
Response to minor comment 16: All denominators have been adjusted for percentage frequencies of cell populations relative to Ter119-CD45- numbers for consistency throughout the manuscript. This has not changed any of the results.

Minor 17. Figure 4c & Extended Figure 5a. It would be better to explain what the track plots is for readers not familiar with this type of plot. More better, it may be consolidated with Figure4b by showing on the x axis the marker genes for Figure 4c.

Response to minor comment 17: We have reanalyzed single cell data and this time included dotplots and violin plots rather than tracksplot for greater clarity.

Minor 18. Figure 5b. The figure legend is not entirely clear in what the connections between the ligand expressing cell subpopulation and the receptor expressing cell population entail. The unit for numbers in each rounded boxes are also not described in the figure legend, noris it described in the method section clearly.

Response to minor comment 18: The values reflect the normalized gene expression based on GEXC calculations as described above (Fig. 3m, comment 4). Lines between populations delineate normalized expression of signaling ligand and cognate receptors and predicted signaling relationships in respective SSC lineage cells to specific hematopoietic lineage cells. The heatmap scheme has been adjusted to Fig. 3e and these details have been added to the figure legend.



Minor 19. Figure 6b. Rather than placing colors to the dots, it would be much better to put another x axis label with the designated experimental condition to aid the readers.

Response to minor comment 19: All graphs now include an additional x-axis information denoting treatment group (Fig. 4c-g).

Minor 20. Figure 6d. The point shapes are all messed up, and the reviewer has no ability to distinguish which group corresponds to the top line (hexagon).

Response to minor comment 20: We apologize for the confusion and have revised the figure for clarity. (Fig. 4c).

Minor 21. Extended Figure 2. CD51 == ITAGV but this is not clearly stated in the figure legend.

Response to minor comment 21: ITGAV has now been replaced with CD51 throughout the manuscript.

Minor 22. Extended Figure 3d. It is hard to tell any difference between 2-month panel and 24-month panel, although the text (Line 132-133) suggests that only 24-month SSCs do not undergo adipogenic differentiation.

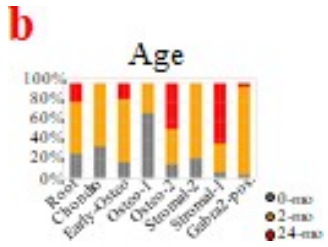
Response to minor comment 22: We apologize for the confusion. Neither 2-month nor 24-month SSCs undergo adipogenic differentiation. We have now clarified this in the text: “Notably, we also observed that neither young nor aged SSCs undergo adipogenic differentiation, even under highly adipogenesis-inducing conditions in vitro.” (lines 143-145)

Minor 23. Extended Figure 4e. The color scheme is reversed (red minus, blue plus), contra to Extended Figure 6a, Figure 5a color scheme. This has to be fixed.

Response to minor comment 23: We thank the reviewer for spotting the error. This has now been corrected.

Minor 24. Extended Figure 5b. As mentioned above, this seems to be the proper leiden clustering results for Figure 4a. It would be more informative for readers though to present the actual frequency (in terms of percentage or better, the actual numbers) of each age groupfor the clusters, preferably as stacked box plots or similar graphical representation rather than UMAP plot.

Response to minor comment 24: We have included a stack plot showing contribution of each age group to the different Leiden clusters (Fig. 3b).



Minor 25. Extended Figure 6k. Inclusion of human specimen (Extended Figure 6k), is cursory, lacking functional characterization of the SSCs in adults. At least either provide similarity in transcriptome (since this is a microarray study and bulk, single cell RNA-seq as well as microarray datasets are presented in this manuscript) to suggest conserved signaling axis (Line 258-259) between mice and humans, or leave this out for future discussion.

Response to minor comment 25: We agree with the suggestion. To avoid controversy stemming from the limited information on a conserved mode of action in humans, we have removed the human data from the updated manuscript.

Minor 26. Extended Figure 7d,e. It appears that the BV/TV is highly variable in Csf1-KO animals, and in wildtype the mechanical strength, BMD, Trab. Th. all have sexual dimorphism raising concerns whether the statistical comparison is properly powered. The reviewer is curious how the equal-variance assumption has been met to conduct unpaired student's t-tests to derive the p-values.

Response to minor comment 26: All results have now been updated to provide proper statistical testing accounting for normality and equal variances. Again, due to limited availability of aged KO mice we were not able to conduct experiments on more animals for Extended Data

Figure 7d-g. Nevertheless, we believe that with the updated statistical testing and the fact that the data supports key findings it is appropriate to present it as is. We acknowledge potential sexual dimorphic factors on effect size (stronger in females, in line with hormone related bone loss) which need to be studied in more detail in future studies, however, the primary observations on trabecular parameters are reflected in the same trends in male and female mice (Fig.3o and Extended Data Fig. 7d-f).

Referee #2 (Remarks to the Author):

1. The manuscript entitled “Aged skeletal stem cells generate an inflammatory niche…” describes an impressive set of experiments demonstrating the age-dependent properties of a population defined by the authors as skeletal stem cells. The paper explores these cells’ molecular properties, regeneration potential and interaction with the hematopoietic system using a series of powerful and well-designed experiments. First, the authors demonstrate aged bones regeneration to be involving a smaller number of distinct clones (and overall much smaller SC output) than observed in younger mice. Then, the authors beautifully dissect intrinsic and systematic properties of this age-related effect, using in-vitro assay, renal grafts and parabiosis. The conclusions from this part of the paper are that aged bones are intrinsically impaired in their regeneration potential, possibly due to some lower potential for stem cell expansion. This part of the paper is excellent in my mind, but there are some questions regarding the linkage between clonaity and ageing phenotype that must be better explained and discussed as detailed below. The authors then try to characterize the molecular properties of the aged-SSC population using scRNA-seq and some additional bulk-RNA-seq, giving rise to some candidate differential markers among which is Csf1. Csf1 function in SSC and bone-regeneration is the focus of the reminder of the paper. The authors convincingly demonstrate Csf1 is overexpressed in aged healing bones, and that Csf1 is affecting negatively healing in-vitro. They also show mice that are haplodeficient for Csf1 had stronger femora which is consistent with a negative role for high Csf1 in maintaining bone structure. But following fracture, Csf1+/- bones regenerate poorly, indicating possible complex feedback and dose-dependent effects are modulating these phenotypes. Finally, the authors demonstrate that that bmp2 treatment of aged mice fractures is improved by adding low Csf1 dose. The second part of the paper uses a powerful array of techniques, and end up with encouraging demonstration of potential treatment, but the basis for some of the arguments is weaker. In particular, it will be important to provide better delineation of the different possible sub-population of Csf1 expressing or non-expressing SSCs – both in young and aged mice as suggested below. It is also important to get more clarity and robustness regarding the effect of adding low Csf1 to the already effective Bmp treatment, and ideally examine the effect this may have on young bones.

   In conclusion, however, I think this is an excellent paper, well written and showing intriguing new results on a very timely and important question.

Response to comment 1: We gratefully acknowledge this Reviewer’s very positive comments and supportive remarks.

Major comments:

2. Clonality and ageing in the bone. The data in figure 1 is very interesting, but it is difficult to link the age related clonality to the other observations in the paper. The authors should provide some clarity about 2 conflicting hypotheses: a) that aged bones are just losing stem cells with age, resulting in less clones, less regeneration, but per single SSC a similar

behavior (e.g. as may be suggested by figure 1 but not the rest of the paper) b) that the pool of bone SSC is ageing “synchronously” such that collectively their potential for regeneration is lower (as suggested by Figure 4-5). Do the authors suggest that both the total number of SSC is going down with age, and their regeneration potential as a group is also decreasing?Is there quantitative evidence that can prove this? (ie. The size of a typical clone in the Rainbow assay is stable with age?)

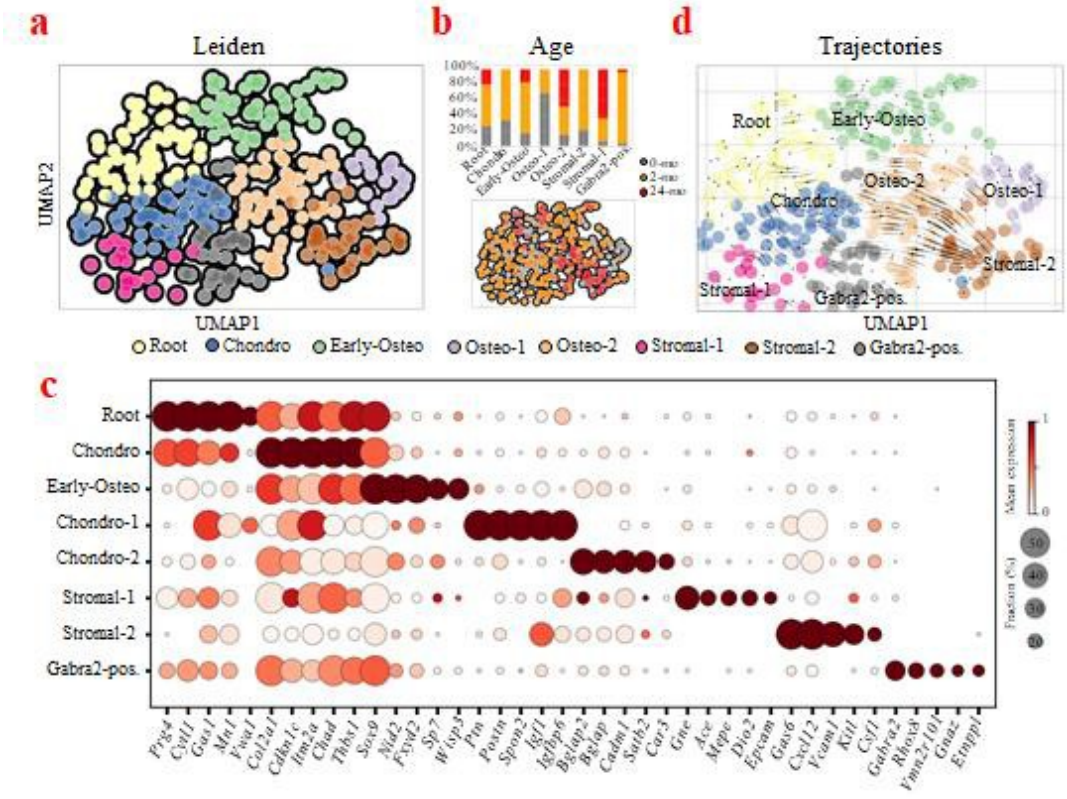
Response to comment 2: We thank the reviewer for these thoughtful questions. Our data supports both mechanisms (a and b) as a possibility. We find that both the total number of SSCs is reduced with age, and their regeneration potential as a group is decreased. First, we find that total numbers and frequency of SSCs/BCSPs are significantly reduced with age during homeostasis and fracture repair (Fig. 1e-h). If seeded in vitro or transplanted in vivo atequal numbers, differentiation potential is clearly reduced in aged versus young SSCs (Fig. 1l-n and Extended Data Fig. 3e-f). The clonal size upon injury (a potent activator of SSCs) isalso diminished with age (Fig. 1a-d and Extended Data Fig. 3b-c). The question if all cells age synchronously is hard to answer and should be a key focus of future research with appropriate genetic and molecular tools but it is beyond the scope of this work at present.

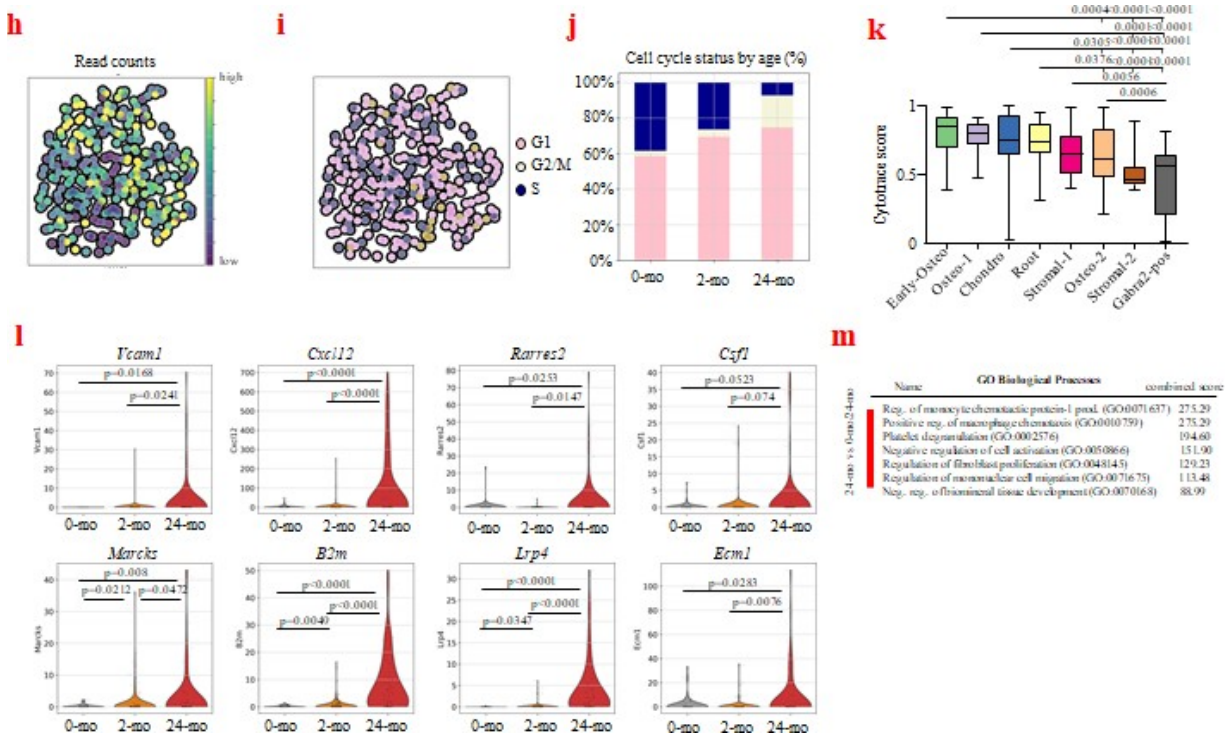
3. In age related clonal hematopoiesis, it is frequently suggested that some recurrent mutation are driving the take-over of clones over HSC niches. It should be clarified if the authors wish to argue a similar process is occurring in the bone or else the belief is that senescence/epigenetic deterioration is the driving force. This is assuming the authors have excluded the possibility of homeostasis with the aged hematopoietic system as a major driver. I find it difficult to understand a prolonged process of ageing as involving homogeneous transition from state A (young) to state B (old). A much more intuitive model is of a mixture of states (A and B) and shifted balance between the two, as sometime argued to occur in the HSC/MPP/GMP/CMP/CLP differentiation hierarchy.

Response to comment 3: As mentioned in response to comment 1 above at this point the mechanism of physiologic aging over time for many stem cell types, including HSCs is still unknown. Somatic mutations, epigenetic alterations, clonal selection, telomerase shortening have all been proposed as possible factors in HSC aging. We have found that aged SSCs are intrinsically different from young and there was little evidence of rejuvenation by exposureto signaling environments in young adults either by heterochronic transplantation or heterochronic parabiosis. We do not yet know if they reach that state by chronic exposure to aged environments over time or if that might affect their epigenetic landscape. Alternatively, some SSCs with reduced osteo-lineage potential could become more prevalent than others due to selection or differences in attrition rates. We have tempered these possibilities throughout the manuscript accordingly. Beyond the scope of this manuscript, we are now identifying novel SSC specific single markers that can be used to generate transgenic reporter mice which could be used to address mechanisms of SSC aging.

4. The analysis of gene expression analysis is somewhat shallow, and its display and explanation in the Methods section is lacking. Figure 4a shows heterogeneity between cells, which is not investigated by the authors. It is only briefly explained by saying that the aged SSC become more homogeneous, which is a very partial explanation. Judging from the figure, it is possible that there is a shift between SSC states during aging, which may provide more insight into the functional decline described in the paper. The only analysis performed on the single-cell data is for differential gene expression, which could have also been performed on bulk data.

Response to comment 4: We thank the reviewer for the comment. We have now re- analyzed our single cell data to more accurately support our findings (Fig 3 and Extended Data Fig. 5). The new analysis revealed that SSCs of all age-groups show striking transcriptomic similarity in their most highly expressed genes. However, using Leiden clustering and trajectory inference analysis we could identify subtle differences in 24-mo SSCs that display a bias towards pro-hematopoietic/inflammatory and anti-chondrogenic lineage directions. In young SSCs on the other hand gene expression reveals higher developmental potential and skeletogenic characteristics (Fig. 3a-d and Extended Data Fig. 5). These findings closely correspond to functional differences in the skeletogenic potential between young and aged SSCs.





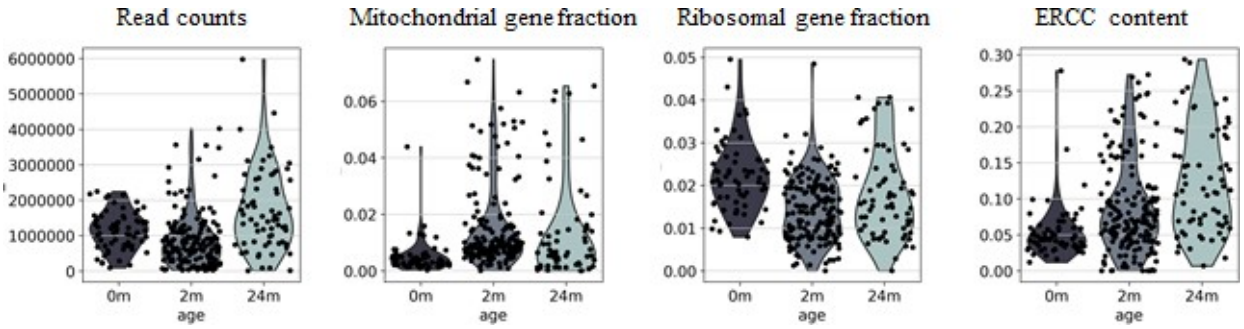
5. It is suggested that the authors will re-analyze their scRNA-seq data to try and characterize variation within and between group, and in particular support their Csf1 analysis downstream. At least the following should be done and reported:

- show QC metric for the single cells – percent mitochondrial, ribosomal and stress (heat shock, proteasome) transcripts per cell, grouped by age. For example, the authors interpret the genes c-Jun, Jund and Junb as associated with commitment to osteochondrogenic fates. These genes are involved in many stress-related processes, often associated also with Irf7/Interferon response and more. One must make sure the young scRNA-seq data is not simply more affected by stress. In that respect it is essential to show batch information for scRNA-seq and make sure it is controlled for.
- show cell cycle analysis (expression of S-phase and M-phase genes) – to demonstrate proliferation rates of the single cell population.
- present a heatmap of de-novo single cell clusters and show the distribution of expression of the genes best separating the clusters. Label cells in each cluster by their age. Perhaps this can lead the authors to some new insight into the regulation of e.g. Csf1 in young mice, or to the existence of “aged” like SSC in younger time points.

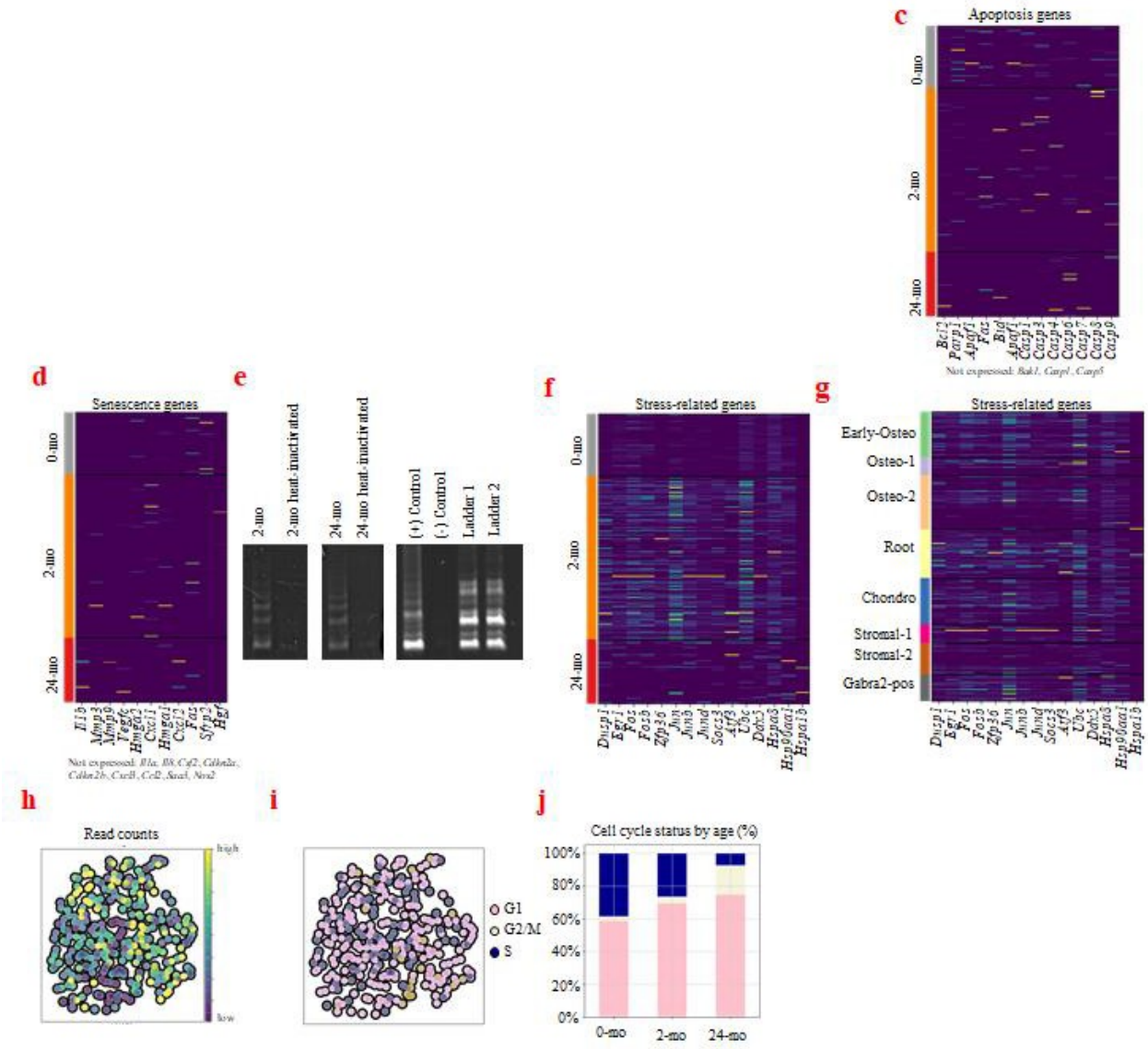
Naturally the above analysis would benefit from generating additional scRNA-seq data. But I do not suggest this should be mandatory. Proper analysis can already improve the results.

Response to comment 5: According to the reviewer’s suggestion, we have included a detailed description of our QC metrics and filtering criteria which are now described in the

method section (lines 991-1066). We have used stringent quality control (QC) parameters to only include high quality single cell data.



Additionally, we have performed batch correction for differences between processed single cell plates. We found that transcriptomic differences between single SSCs from 0-mo, 2-mo,and 24-mo mice are subtle. Varying the number of principal components (PCs) for our clustering analysis did not significantly change cell separation. Based on the PC elbow plot we settled on 36 PCs by which we were able to identify specific differences between SSCs of different age groups (Fig. 3a-d & Ext. Data Fig. 5). Extended Data Fig. 5c-j also shows thatcell cycle status, read counts, and expression of stress-, senescence-, and apoptosis-relatedgenes did not affect clustering of single cells.

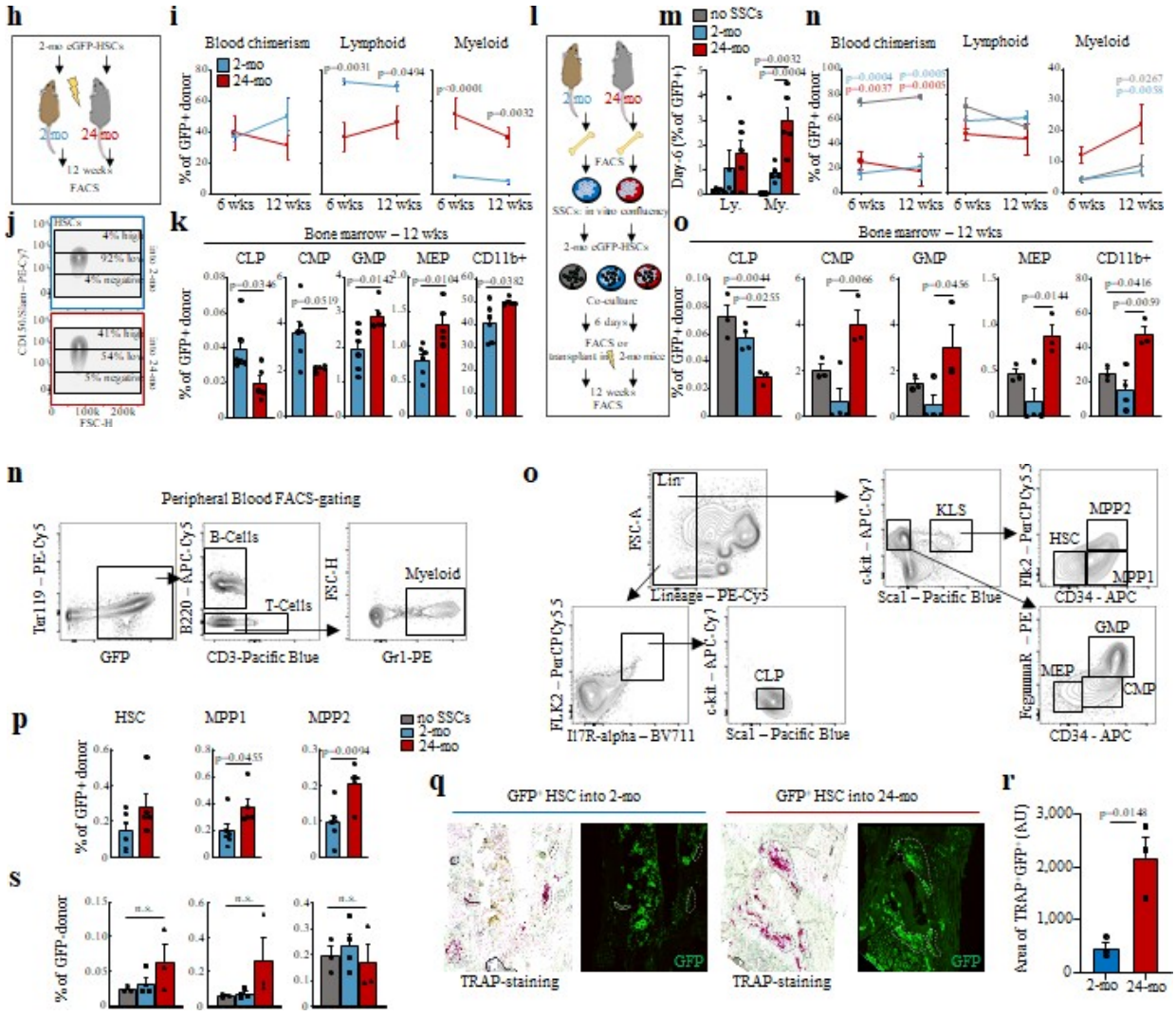


New data also includes more information on Leiden cluster composition and specific differential expression of genes between clusters and age groups as shown in response to comment 3 (Fig. 3b-c & Extended Data Fig. 5l-m, Supplemental Information 1).

6. It is not clear how the expression values in Figures 4e-f (and similarly in Extended Data Figure 5c-d) are calculated, and they appear to exaggerate differential expression. For example, Figure 4b shows a very subtle difference in Csf1 expression, while 4f shows a very distinct one. The authors should instead show boxplots or dotplots for key markers, with careful statistics and normalization to library depth and QC as discussed above.

One possible approach for more sensitive analysis is to search for SSC genes that are correlated with Csf1 levels within each age group. It is not a-priori clear that top correlated genes would be those that are also differentially expressed in young/aged mice.

Response to comment 6: We thank the reviewer for the comment and have now presented single cell gene expression as dotplots and violin plots to provide more clarity. The transcriptomic signature of single aged SSCs skews from younger age groups by expressing pro-hematopoietic/-myeloid cytokines which is further supported by additional bulk transcriptomic and functional data (Fig. 3e, m). In additional new experiments, we demonstrate a direct connection between SSC-derived myelo-supportive stromal lineages and myeloid-shifted HSCs (hematopoietic stem cell) output in vitro and in vivo upon co-culturewith SSC derived stroma (Fig. 2h-o, Extended Data Fig. 4n-r).



Finally, genetic and in vivo rescue experiments support upregulated expression of Csf1 as a key driver of functional impairments in the aging skeletal system (Fig. 3p-r and 4).

7. The Methods section should provide more detail on how the data were analyzed, and the code for the analysis should be made publicly available (before publication).

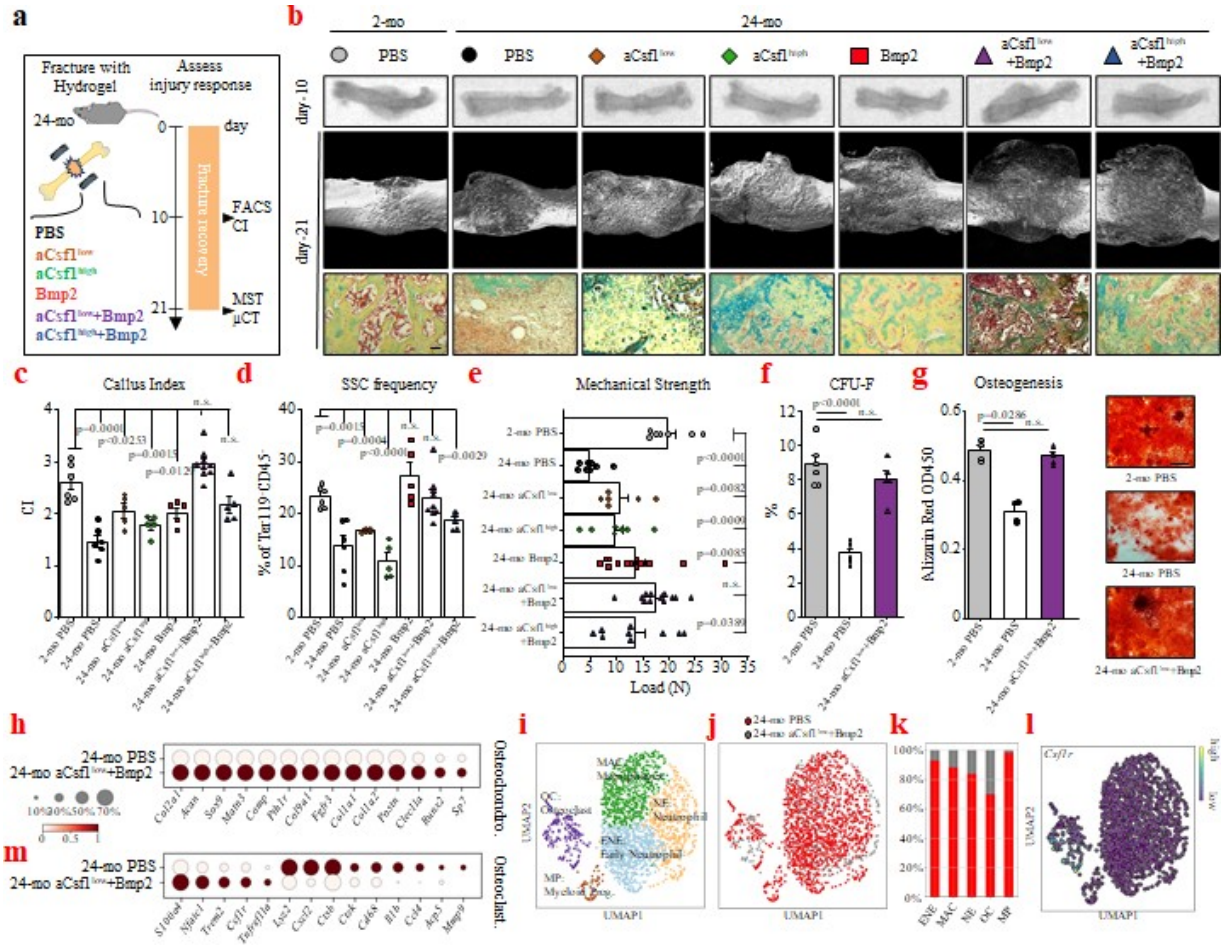
Response to comment 7: The Method section has been updated (lines 991-1066) and the processing notebook is available as described in the method section and will be further included in the GEO submission.

8. The authors link the bias of aged SSC to Thy1+ stroma to the shift of HSCs to the myeloid lineage in aging (e.g. "our findings indicate that the aged SSC lineage favors the selective expansion of myeloid-skewed HSCs"). These conclusions are not adequately supported in the paper. Even if previous work showed that Thy1+ supports myeloid cells, and this work shows that SSCs create more Thy1+ cells in aging, it is not shown that this increase leads to more myeloid cells in aging. The only evidence presented here for this conclusion is the upregulation of csf1, the concomitant csf1r by myeloids, and the GO enrichment of "macrophage migration", and this evidence is largely anecdotal. Other evidence for the effect of SSC on hematopoiesis, such as the high expression of Cxcl12, Kitl and Vcam1 in aging SCCs is also not convincing. These results are interesting and can be briefly mentioned, but no strong conclusions can be drawn from them.

Response to comment 8: As pointed out by the reviewer, aged SSCs show a distinctly altered lineage output in vitro and in vivo (Fig. 1k and Extended Data Fig. 2e-f). We have nowconducted additional experiments showing (see comment 4 response above) that exposure to an aged SSC-derived stromal environment functionally skew young HSC towards myelopoiesis (Fig. 2h-k and Extended Data Fig. 4n-p). We first find that young HSCs transplanted into aged mice are myeloid skewed and supportive of enhanced osteoclastogenesis. We then show in co-culture experiments that stroma derived from agedSSC also skew young HSCs towards myelopoiesis (Fig. 2l-o and Extended Data Fig. 4q-s). Together these new studies support our previous findings suggesting that aged SSC derivedlineage potentiates the selective expansion of myeloid-skewed lineages from HSC.

9. The combinatorial treatment data is really interesting, but the high variance in the overall effect (Fig 6e) makes it difficult to argue convincingly that Bmp2 treatment is improving when adding aCsf1low. Statistics must be improved here. One can suggest (but I don't suggest this should be required) to perform some additional follow ups (for example gene expression of Bmp2 treated SSCs with and without Csf1low and Csf1high). The authors are heavily relying on the results to motivate their story – it is important to make sure the result is robust.

Response to comment 9: We have added additional animals for the rescue experiments and included a young control group as a more accurate reference point. These additional measures have strengthened our results. We have also conducted 10X scRNAseq on aged PBS-control and aged aCsf1/rBmp2 treated groups showing transcriptional rejuvenation of fracture healing in treated aged fractures vs untreated fractures (Fig. 4).



10. Typos / small errors:

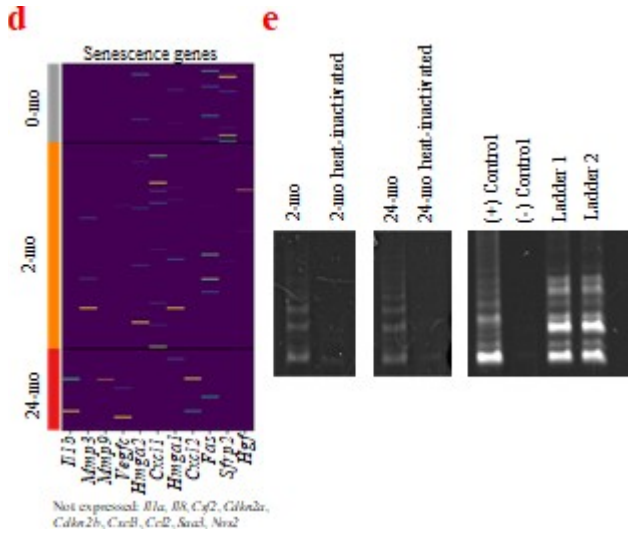
- According to the legend of Figure 4a, the figure shows Leiden clustering of the cells, but the cells are marked by the mouse age.
- "SSCs could also contributes" – should be "contribute" (page 15).

Response to comment 10: Thank you! Corrected.

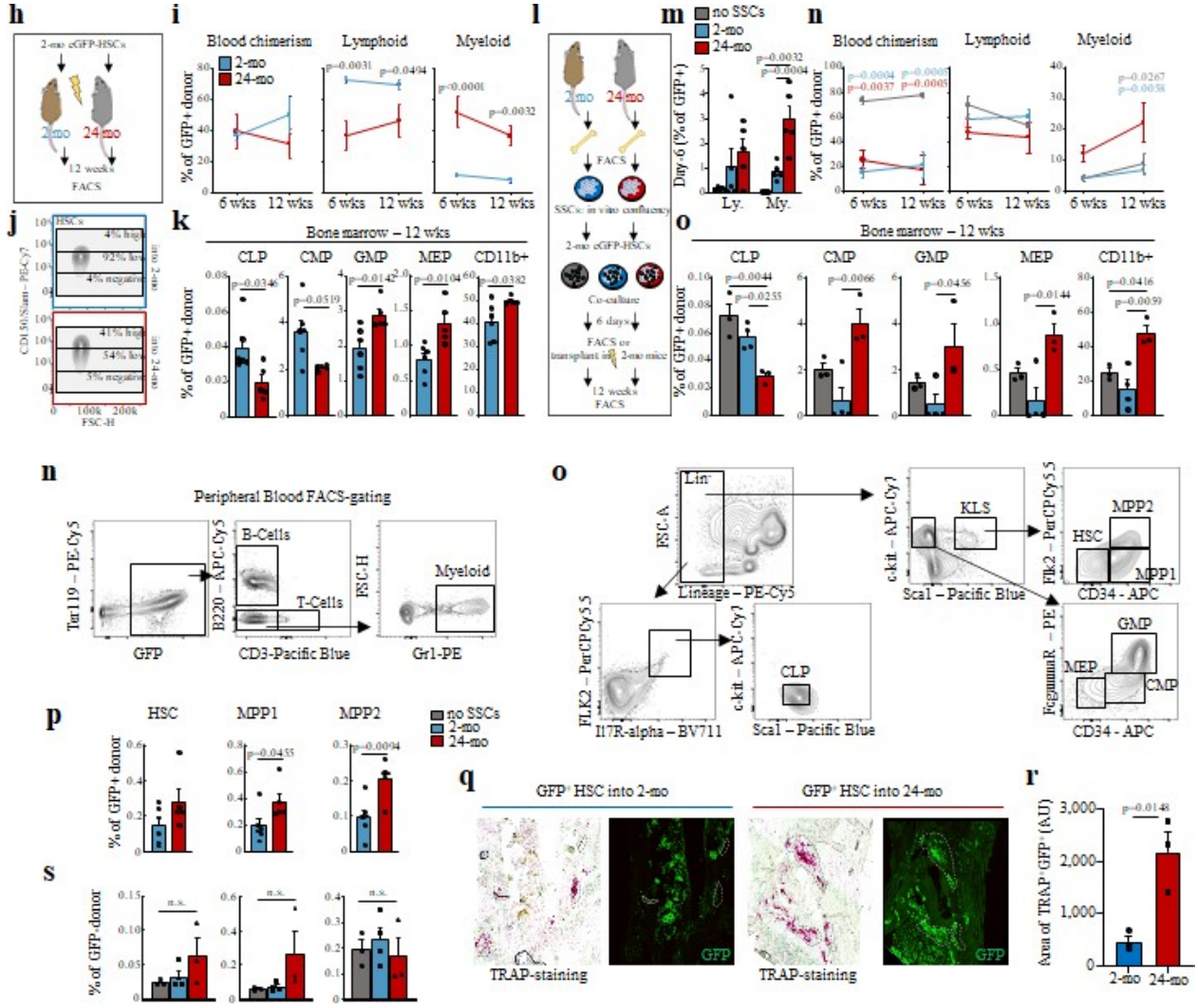
Referee #3 (Remarks to the Author):

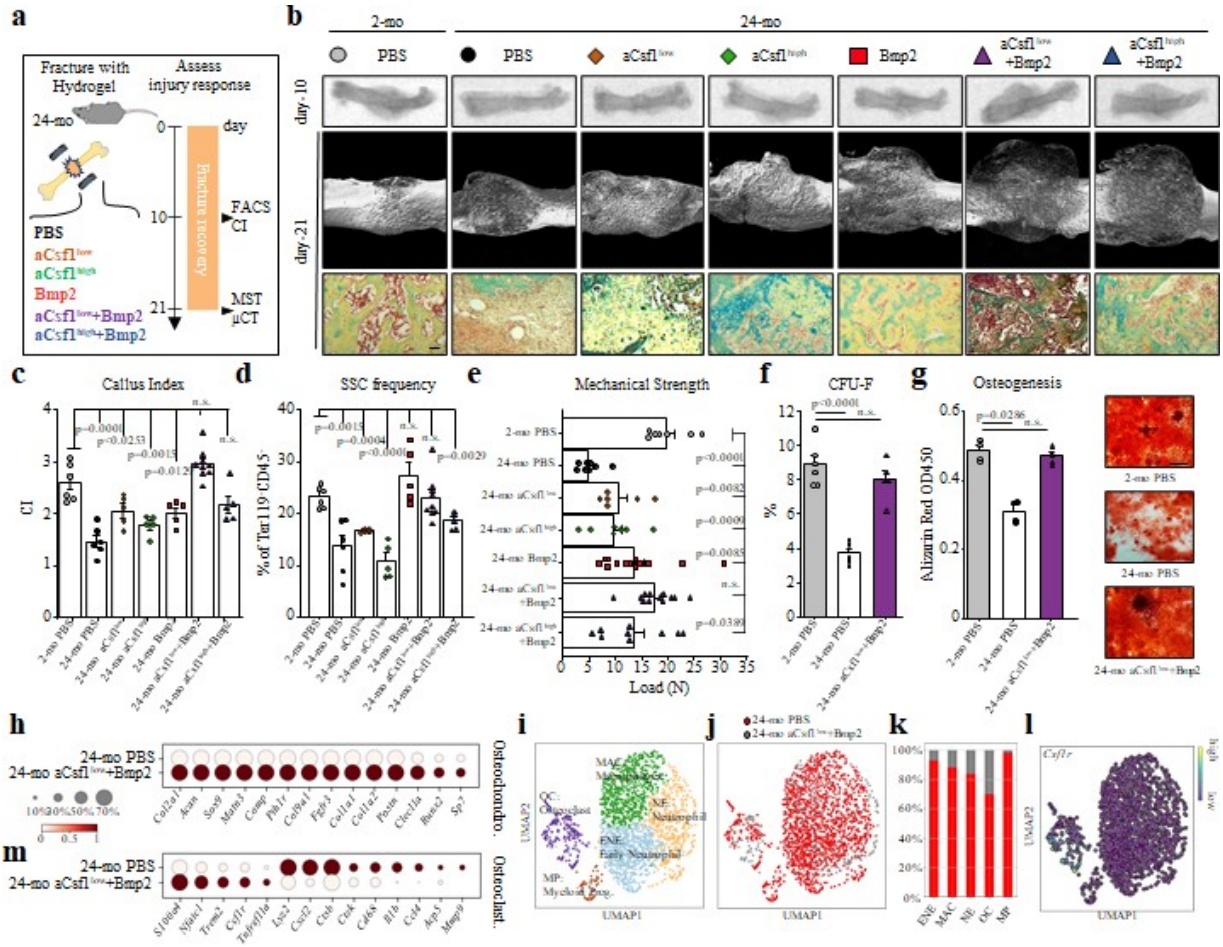
1. In this manuscript Ambrosi et al. evaluate whether aging associated decreases in bone mass are due to SSC intrinsic pathology and evaluate the mechanisms involved. The topic addressed is both novel and important, as while there have been extensive studies of mechanisms involved with aging-associated skeletal pathologies, none of this prior work have resolved the basis of aging associated skeletal pathologies down to discrete mesenchymal populations. However, relatively little data is offered that directly maps aging associated cellular dysfunction to defects in SSCs.

Response to comment 1: We thank Reviewer 3 for the positive and supportive remarks. In previous studies, SSCs have been shown to be the major stem cell type giving rise to mature bone, cartilage, hematopoietic supportive stromal tissues (Chan et al. 2015, PMID:25594184; Worthley et al. 2015 PMID:25594183). SSC activity is essential for both skeletal formation and regeneration as shown by experiments using genetic ablation and by ablation through irradiation. Additional studies by other investigators were also consistent with the earlier findings (Debnath et al 2018, PMID: 30250253; Mizuhashi et al 2018, PMID: 30401834). Here we show that aging significantly depletes SSC frequency while dramatically reducing their skeletogenic potential. Per the reviewer's remark, we directly tested for aging associated cellular dysfunction to defects in SSCs. We show in Figure 1 that SSC frequency, clonogenicity, skeletogenic potential(in vivo and in vitro), and injury response (in vivo) all decline in 24-mo as compared to 2-mo mice. Transcriptomic, genetic, and functional experiments provide additional mechanistic characterization (Figures 3-4). We were actually surprised that some traditionally considered mechanisms for cell aging such as cellular senescence/reduced telomerase activity did not seemto be involved in SSC aging (Extended Data Fig. 5d-e).



Rather, it seems SSC aging likely involves epigenetic mechanisms that change the transcriptionof key genes involved in skeletal lineage commitment. Additional experiments are also now included to better support previously presented evidence as requested by the reviewers. (Fig. 2h-o and 4).


h
i
l
m
n
j
k
o
Blood chimerism
Lymphoid
Myeloid
2-mo
24-mo
% of GFP+ donor
6 wks
12 wks
no SSCs
Day-6 (% of GFP+)
Ly.
My.
FACS
SSCs: in vitro confluency
2-mo cGFP-HSCs
Co-culture
6 days
FACS or transplant into 2-mo mice
12 weeks
Bone marrow – 12 wks
CLP
CMP
GMP
MEP
CD11b+
HSCs
4% high
92% low
4% negative
41% high
54% low
5% negative
into 2-mo
into 24-mo
CD150/Slam – PE-Cy7
FSC-H
100k
200k
n
o
Peripheral Blood FACS-gating
Ter119 – PE-Cy5
GFP
B220 – APC-Cy5
B-Cells
T-Cells
CD3-Pacific Blue
FSC-H
Myeloid
Gr1-PE
Lin-
FSC-A
Lineage – PE-Cy5
c-kit – APC-Cy7
KLS
Sca1 – Pacific Blue
Flk2 – PerCPCy5.5
MPP2
HSC
MPP1
CD34 - APC
FLK2 – PerCPCy5.5
IL7R-alpha – BV711
CLP
FcgammaR – PE
GMP
MEP
CMP
p
s
HSC
MPP1
MPP2
no SSCs
2-mo
24-mo
% of GFP+ donor
% of GFP-donor
n.s.
q
r
GFP+ HSC into 2-mo
GFP+ HSC into 24-mo
TRAP-staining
GFP
Area of TRAP+ GFP+ (AU)
3,000
2,000
1,000
0
2-mo
24-mo



2. A similar concern is that th**e primary mechanism identified, increased CSF1 secretion,**is a well known regulator of osteoclastogenesis and that the experiments performed cannot distinguish a role of CSF1 in SSC related mechanisms of skeletal aging versus the general and known pro-osteoclastogenic functions of CSF1.

Response to comment 2: We apologize for the confusion. There are two key SSC-related dysfunctions, one being in bone formation and the second in promotion of osteoclastformation in part through generation of stromal cells that express high levels of Csf1. While Csf1 is a known pro-osteoclastogenic factor the increased expression and secretion of Csf1by aged SSC-derived cells establishes a novel finding and connects a known mechanism thatdrives bone loss to the impaired properties of SSCs. Furthermore, we do not claim that SSCderived Csf1 is the only mechanism driving age related alteration, but also demonstrate thatSSC aging is involved in skewing of HSC towards a general pro-inflammatory, pro-myeloid lineage commitment that are involved in aging phenotypes in a variety of tissues (Fig. 2h-o, 3).

While the methods used here are technically sound, there are several major concerns that the approaches applied do not appear appropriate to address the questions posed or are unable to adequately support the conclusions proposed. Specific details are described below.

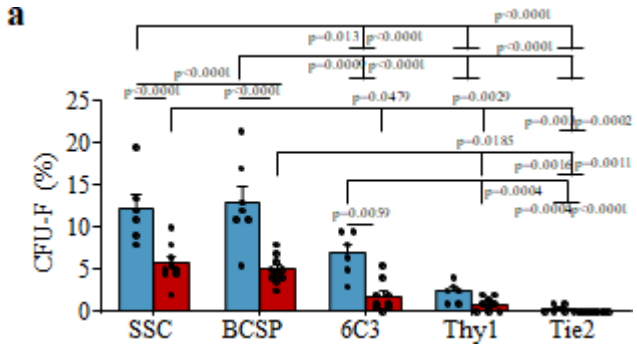
Major points

3. Prior reports on human and murine SSCs identify these cells as growth plate resident cells. However, here the properties of these cells are being studied under aging conditions where the growth plate is "lost", especially in humans where there is essentially complete growth plate fusion. This apparent conflict requires clarification.

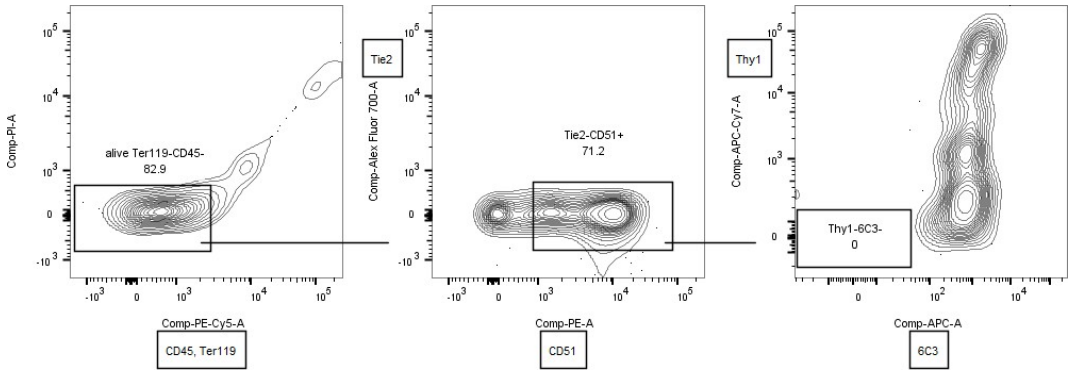
Response to comment 3: We initially identified mSSCs as growth plate resident stem cells (Chan et al 2015, PMID: 30241615) but since then have shown that mSSCs are abundant in periosteal, cartilaginous and bone marrow regions of postnatal long bones (Murphy et al 2020, PMID: 32807933), which has also been independently confirmed by others (Debnath et al 2018, PMID: 30250253; Mizuhashi et al 2018, PMID: 30401834).

4. While the data on transplantation of SSCs and BCSPs post fracture to evaluate the size and composition of the in vivo ossicles they form in a transplantation system is important, it cannot be excluded that this result reflects an attenuation of the ability of non-stem progenitors generated from aged SSCs to progress to mature cell fates. The finding that BCSPs also display a similar attenuation of ossicle formation capacity and in vitro differentiation capacity raises the possibility that some of the effects seen occur in a "post-SSC" cellular compartment or reflect a broad and general dysfunction of mesenchymal cells not specific for SSCs. -

Response to comment 4: We thank the reviewer for the comment. Since SSCs and BCSPs give rise to downstream progenitor populations of bone and cartilage cell types of the skeleton, the loss of functional activity at the apex of the lineage tree (SSCs/BCSPs) is very likely the major contributor to age-related skeletal impairments. Phenotypic SSCs/BCSPs remain the population with the highest clonogenic potential throughout mouse life and in aged bones (Extended Data Fig. 3a).



This is consistent with our previous findings showing that purified cell populations other than SSCs and BCSPs do not contribute to multi-lineage output in the transplant setting (Chan etal 2013 and 2015, PMID: 26216955 and 25594184), while at the same time cell properties and lineage output are altered at the stem cell level in aged mice (Fig. 1k-n, Extended Data Fig. 2e-f). Of note, SSC/BCSP depleted cells do not give rise to SSC/BCSPs upon 7-days invitro culture:



Altogether, our data strongly suggests that the major age-related bone-forming deficiencies originate at the SSC level but this does not exclude the possibility that downstream populations derived from SSCs also inherit age-related stem cell defects. Indeed, our data shows that SSC-derived stromal cells also contribute to skeletal dysfunctions in aged bones by promoting myeloid-skewed hematopoiesis and higher levels of osteoclasts (Figure 2&3). Future work is needed to determine if downstream cell types inherit functional genetic and epigenetic changes from aged SSCs that contribute to the aging phenotype.

5. The data currently offered does not clearly deconvolute the relative contribution of these SSC versus "post-SSC" mechanisms. An important aspect of performing this deconvolution is to address the degree to which aging specifically influences SSC stemness properties, such as by assessing relative self-renewal capacity of aged versus young SSCs in a competitive or limiting dilution system.

Response to comment 5: We have conducted additional CFU-F assays now including four defined skeletal lineage populations to supplement in vitro and in vivo differentiation assays on highly purified SSC/BCSP populations (Fig. 1 and Extended Data Fig 2a-c). These are standard assays used to determine the stem cell properties of SSCs. In contrast to HSCs, which can reconstitute hematopoiesis at the single cell level and can migrate to and from bone marrow niches through the circulation, SSCs do not have these properties. Therefore, unfortunately, it does not seem feasible to conduct “competitive or limiting dilution system” experiments. We have however shown that purified SSCs from young and aged mice behave differently under identical conditions in vitro and in vivo (Fig. 1 and Extended Data Fig 2a-c).

6. Along similar lines, the only experiment directly addressing whether aging results in differences in the cellular output specifically of SSCs is conducted in vitro in Extended Fig 2e, which only shows modest differences.

Response to comment 6: We kindly disagree. We have previously described the SSC lineage tree and show flow cytometric analysis of abundance of the different cell types derivedfrom young and SSCs in vitro as well as upon activation due to fracture injury in vivo (Fig. 1kand Extended Data Fig. 2e-f). The differences in lineage output at the stem cell level can alsobe deduced from single cell RNA-sequencing and functional assays showing clear and significant differences between young and aged groups for differentiation as well as during rescue experiments (Fig. 1, 3 and 4).

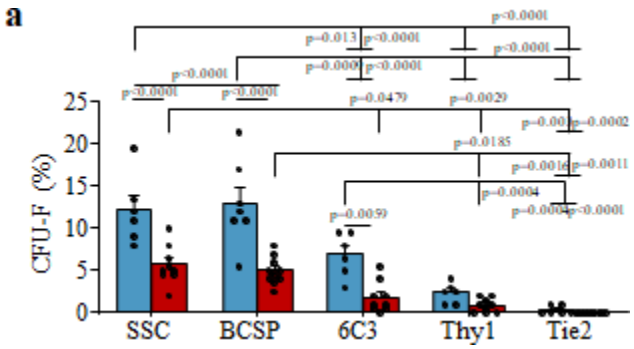
7. In vitro differentiation is well known to not necessarily reflect in vivo differentiation capacityin bone. It will be important to specifically assess the cellular output of aged vs young SSCs versus BCSPs and pre-BCSPs in vivo to address the stem cell specific nature of aging induced dysfunction. This data is needed to justify the claims made in the section heading (line 78-9).

Response to comment 7: This has been addressed by transplanting either SSCs or BCSPs underneath the renal capsule. Results show that aged BCSPs show corresponding defects to SSCs (Fig. 1m-n and Extended Data Fig. 3e-f).

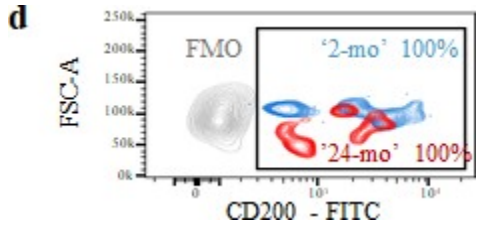
8. A distinct but related issue is that the SSCs examined at older ages may represent a qualitatively different cell type (as opposed to the interpretation favored in the manuscript that aged SSCs represent the same cells as young SSCs, albeit with quantitative alterations in mature tissue production capacity).

This may simply reflect that the flow panel applied here is not able to resolve the compositional transitions occurring with aging within the SSC compartment. This concern is supported by the scRNA-seq data in Fig 4a, b and Extended data fig 5a, b showing qualitativechanges in makeup of the SSC pool, with multiple clusters present at young ages and only asinger cluster present in older mice. The omission of CD200 from the analysis (see major point 4 below) also raises the level of concern that unappreciated heterogeneity may be central to the observations made here.

Response to comment 8: As already addressed in comments to Reviewer 1 and in responseto this reviewer's comment 4 we have conducted functional assays on different purified downstream populations (BCSP, Thy1+, 6C3+) of the SSC lineage from young and aged mice and show that phenotypic SSCs remain the population most enriched for stem cell characteristics (Extended Data Fig. 3a).

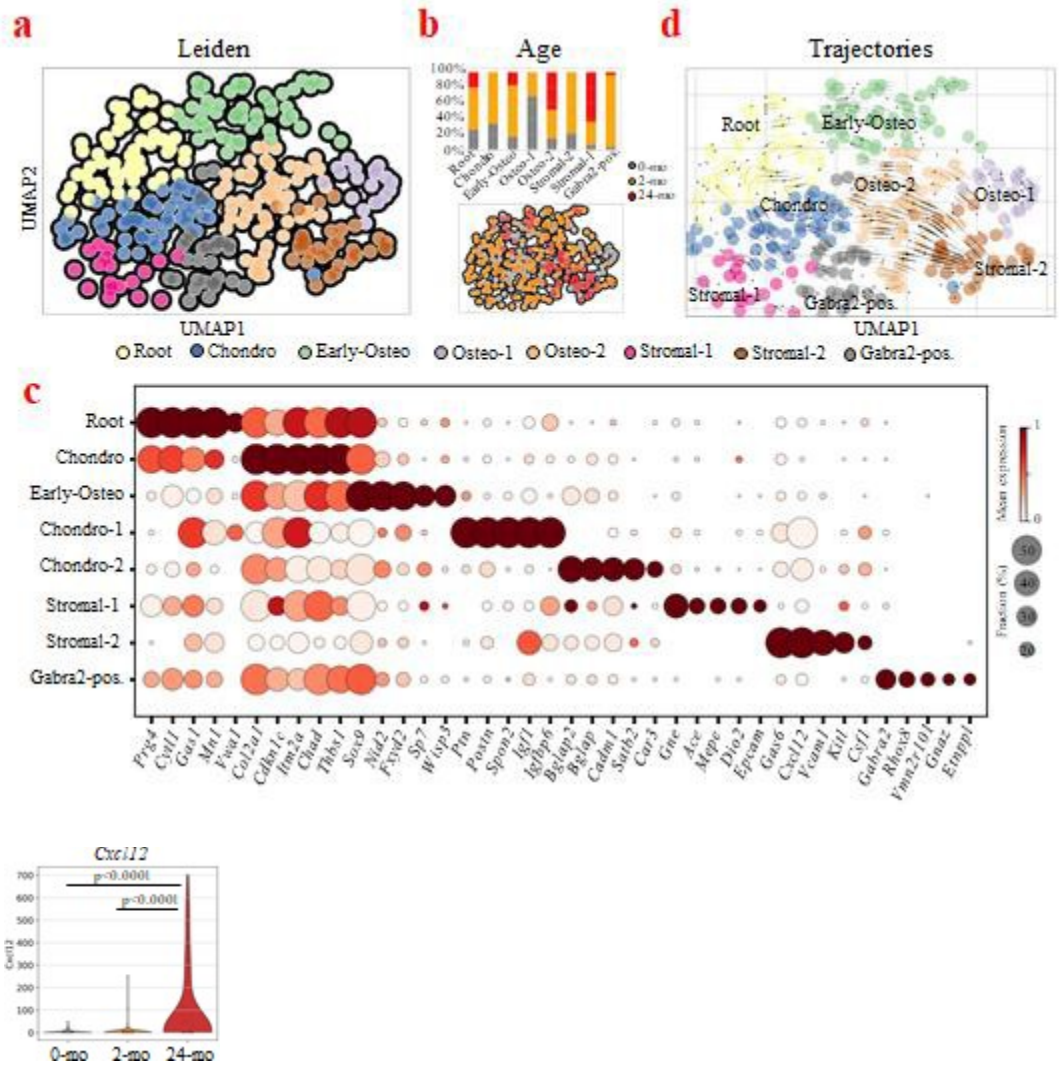


Similarly, CD200 does not further distinguish SSC and pre-BCSP populations in adult mice as shown in new Extended Data Figure 2d.



9. Indeed, it would seem likely the changes in cell types comprising the SSC pool as suggested by thesescRNA seq studies would be central to the aging phenotypes observed. However, these changes in the cell types falling within the SSC pool are not clearly described and their potential significance to the aging associated phenotypes is not examined experimentally. Specifically, finding that Cxcl12 is increased in 24mo SSCs suggests that this pool may be now be comprised of Cxcl12 abundant reticular (CAR)/LEPR+ cells that are normally not present within the Chan et al. murine SSC definition, as CAR cells are marrow stromal elements separate from cell populations in the growth plate region.

Response to comment 9: We have reanalyzed our scRNAseq data to address this issue more precisely (Fig. 3a-d & Extended Data Fig. 5l). Cxcl12 is present in younger SSCs and increases expression with age.



Looking at cluster specific gene expression and results of trajectory inference we conclude that young SSCs are diverse consisting of potentially pre-primed subpopulations for osteochondrogenic and stromal lineages (Fig. 3a-d and Extended Data Fig. 5). Aged SSCs lose the transcriptomic imprint for chondrogenic lineages but instead increase genes representative of a specific subpopulation of stromal cells.

LepR universally traces all cells of the SSC lineage in young and aged mice and merely enriches for heterogeneous populations with high CFU-F capacity. Similarly, Cxcl12 is highly abundant throughout the bone in various cell populations. This also means that LepR and CXCL12 are not restricted to reticular CAR cells. This reflects the greater specificity of usinga combination of cell surface markers over single genetic labels.

10. In comparison with the prior Chan et al. report on murine SSCs, CD200 has been excluded from the definition used here. This omission is important as it represents a substantial change in the SSC definition. SSCs here are only designated as Lin-CD105-

ITAGV+ cells, a much larger and presumably more heterogenous population than the CD200+ subset of the above. This scheme also omits the previously described “pre-BCSP” category and merges these cells with SSCs. Are CD200+ SSCs present in aged mice or does the pool of cells designated as SSCs here predominately reflect pre-BCSPs in aged mice? This issue has important implications as to whether the defects observed here truly reflect SSC dysfunction.

Response to comment 10: We thank the reviewer for this comment. As described in response to comment 8, the initial identification of the mouse skeletal stem cell indeed utilizedCD200 as an additional criteria to distinguish SSC and pre-BCSP populations. However, thiswork was based on fetal and newborn SSCs (Chan et al. 2015, PMID: 25594184). In adult SSC lineages all CD51+Thy1-6C3-CD105- cells were found to be CD200 positive, thus not allowing further separation by this marker. We now include FACS analysis comparing youngadult (2-mo) and aged (24-mo) mouse bone single cell solution gated for the CD51+Thy1- 6C3-CD105- cell population showing that virtually all adult SSCs express CD200 (Extended Data Fig. 2d, see comment 8). We have additionally conducted CFU-F assays on SSCs andpreviously defined downstream populations to show that phenotypic SSCs and BCSPs maintain the highest clonogenic potential throughout age, i.e. stem cell characteristics(Extended Data Fig. 3a, see comment 8).

11. Dynamic histomorphometry is needed for the experiments in Fig 3b, c to assess the degree to which these results here reflect differences in anabolic versus catabolic capacity. Supplementing this with serum levels of CTX or a similar catabolic marker is helpful to provide an assessment of catabolic activity. This question of anabolic versus catabolic nature of the differences observed is not addressed by the bone marrow transplant study performed in Fig 3j, as differences in catabolic activity can also be intrinsic to alterations in the osteoclastogenic capacity of the mesenchymal compartment or the influence of other non-circulating cell types present in bone on osteoclastogenesis.

Response to comment 11: We show that serum Rankl and Ctx1 levels are reduced in heterochronic aged versus isochronic aged mice supporting the notion that a young circulation can have positive effects on hematopoietic cell types (Fig. 2g and Extended DataFig. 4g-h). However, analysis of local osteoclast activity by TRAP staining at fracture sites inparabionts seems to be less altered (Extended Data Fig. 4i-j). This suggests that the local environment created by SSC-lineage cells (and potentially others) could partially off-set systemic concentrations of cytokines and growth factors.

12. In many of the mouse studies, the "n"s employed seem fairly small given the expected variation in the models employed, such as in Fig 3 at 3-4 per group for many groups. How many independent repeats were performed and how many mice were analyzed across all repeats in total? Review of the total data would be helpful in establishing confidence in these results. Similar comments apply to Fig 5e. It is unclear as presented whether the increase in

BMD comparing HA to IA samples in Fig 3b is biologically meaningful or statisticallysignificant, and the statistics for this comparison should be clarified. The p values provided appear to be comparing the HA and IA samples to the IY samples, but this is likely not the most relevant comparison for these groups. Instead the comparison of greatest interest is comparing HA to IA specimens.

Response to comment 12: We are limited by our institution's IACUC protocol to utilize the minimum number of animals required to reach statistical significance for these experiments.All data points reflect individual mice from at least two separate experiments as described inthe method section. We also would like to clarify that statistical testing between all groups has been done and statistically significant comparisons are annotated with the p value given.We have added bars for groups showing statistically significant changes and amended figure legends to describe the type of statistical testing. We now also provide a reporting summary showing raw values for each figure.

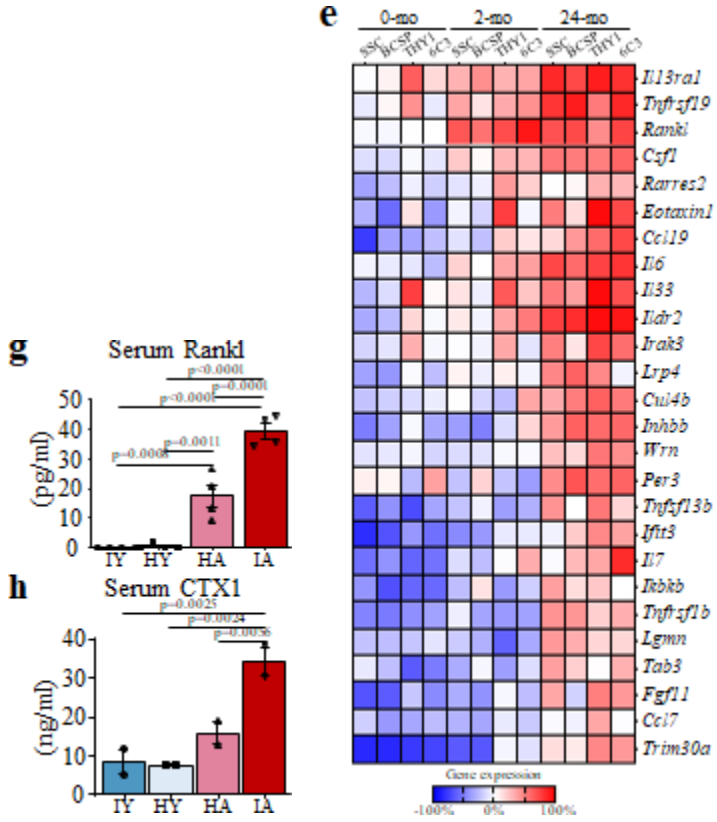
13. Bone mass phenotypes are expressed in many figures as BMD instead of the more standard trabecular BV/TV and cortical thickness. Does the data in Fig 3b and 3j represent uCT or use of a DXA-like platform? Use of uCT is important given that DXA measurements are relatively insensitive and may mask important phenotypes. Cortical and trabecular phenotypes should be separately reported using standard nomenclature for uCT, as there are compartment specific differences in response, and phenotypes in cortical thickness have a more direct relevance to human fracture risk. At least for a subset of key experiments, the fracture phenotypes throughout would be better reported as an analysis of total callus bone volume as determined by uCT as opposed to the 2D radiograph based callus index.

Response to comment 13: All bone parameters (BV/TV, BMD, etc.) are based on microCT measurements of trabecular areas as described in the method section. To our knowledge either BMD or BV/TV are accepted reporting parameters. In most cases BMD and BV/TV show the same trends, similar to what we observe with callus index readouts (Marecic et al 2015 PMID: 26216955; Tevlin et al 2017 PMID: 28077677). We have added cortical parameters for experiments in Extended Data Fig. 1c and 7e. In our experience trabecular bone mass is a superior predictor of bone brittleness in mice. For fracture callus analyses wehave calculated bone parameters based on total callus volume. If possible, we provide callusindex and uCT data. For the experimental results in Extended Data Fig. 4m (formerly Fig. 3j)we were unable to conduct uCT.

14. While there is some nuance to the data, looking at the results in Xiong et al. Nature Communications 2018 in mice with sheddase-resistant RANKL in comparison with mice with complete RANKL or RANK deficiency overall emphasizes that the role of soluble RANKL in bone metabolism is fairly minor. Thus, the lack of differences in soluble RANKL are of unclear relevance to the phenotypes observed and don't necessarily offer evidence that non-RANKL

factors are involved. Measurement of tissue-level RANKL expression would be more relevantin this context.

Response to comment 14: We agree with the reviewer that circulating RANKL is not necessarily an indicator of local bone-resorption. The parabiosis data supports this notion, as noted above (comment 11), showing that local TRAP staining of fracture calluses was not significantly altered in heterochronic aged mice, although circulating RANKL was significantly reduced (Extended Data Fig. 4g-j). Our microarray data of skeletal lineage populations further showed that RANKL is expressed in all cell types in young and aged mice (Fig. 3f) and the overall expression levels between young and old amongst SSC lineage populations were similar.

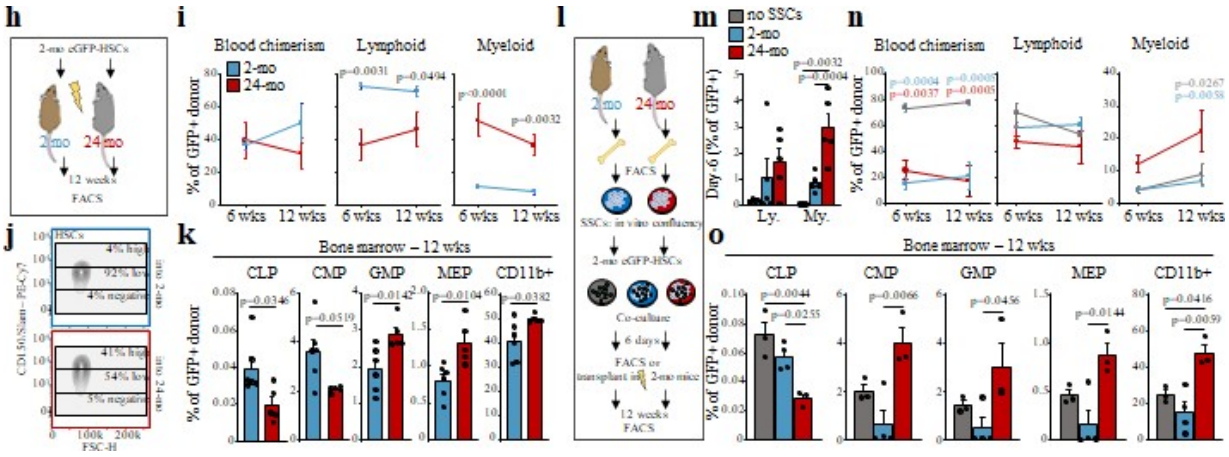


This data of course cannot distinguish between membrane bound and circulating RANKL. Nevertheless, this observation led us to investigate the role of the second necessary factor of osteoclastogenesis, Csf1, in our downstream analysis. Here we do see that expression ofCsf1 is progressively increased with age in SSCs and downstream lineages and we find thatblocking Csf1 specifically can ameliorate bone healing in combination with Bmp2 (Fig. 3&4).

15. The interpretation offered for Fig 3h, that myeloid skewing of HSCs reflects altered stromal signaling/that myeloid skewing with aging is an SSC property that can be rescued in HA parabionts doesn't appear to be supported by the current experiments. It is unclear how it can be excluded that the phenotype seen in the HA vs IA parabionts reflect direct actions of young

parabiont derived circulating factors on HSCs as opposed to the stroma. The system used with transplantation of these HSCs to young recipients would appear to favor the former, as these effects persist post-HSC transplantation. It may be more informative to examine the HAversus HY derived HSCs directly in the context of the bones of either the HA versus HY parabionts to determine to what degree any myeloid skewing phenotypes reflect cell intrinsiceffects of aging on the local environment versus effects mediated by circulating factors.

Response to comment 15: We agree and also hypothesized that circulatory factors from young blood could partially rejuvenate HSC function in HA HSCs. However, in parabiosis it is also possible that young circulating HSCs could migrate to aged bones leading to the appearance that “rejuvenated” HSCs are giving rise to balanced rather than aged myeloid-biased hematopoiesis. To further investigate if an aged stem cell niche skews lineage outputof HSCs, we have transplanted young HSCs into young and aged hosts. The results suggestthat the aged microenvironment could lead to myeloid skewing of HSCs (new Fig. 2 h-j). Young GFP HSCs transplanted to young or aged mice also preferentially gives rise to myeloidlineage osteoclasts (new Extended Figure 4q). To more directly address the role of SSCs and downstream lineage cells we have co-cultured young HSCs with young and old SSC cultures. After six days of co-culture significantly more myeloid lineage cells were present in HSC co-cultures with aged SSCs. If transplanted in a hematopoietic reconstitution setting young HSC-young SSC co-culture cells produced significantly more lymphoid versus myeloid cells compared to aged SSCs (new Fig. 2 k-n). This further supports the hypothesis that the aged local environment, and at least in regards to SSC-derived lineage cells, could induce aged, pro-inflammatory myeloid-skewed hematopoiesis from HSC.



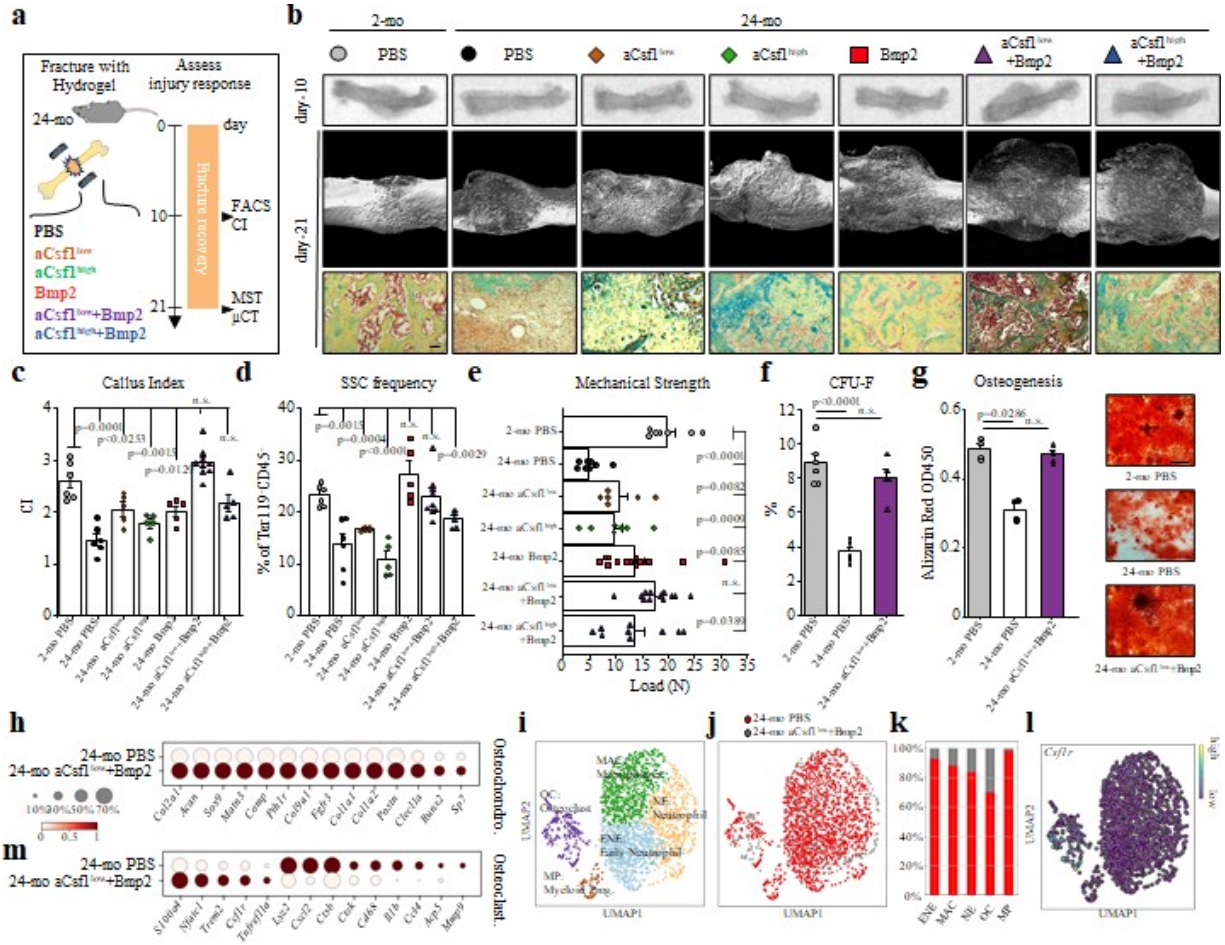
16. The experiment supplementing bone marrow cultures with RANKL and CSF1 (extended data fig 6, lines 261-5) does not show that increases in Csf1 expression and secretion are an important factor contributing to the increased osteoclastogenic capacity of aged mesenchymal cells. Instead, this data shows an increased osteoclastogenic potential of the aged bone marrow stroma, which may be due to any number of factors, including differences in osteoclast precursors, differences in the receptiveness of these precursors to

CSF1/RANKL or differences in the amount of co-stimulatory factors present that enhance osteoclast differentiation. Additionally, no direct evidence is offered to functionally establish that any aging-associated osteoclastogenic phenotypes are specifically attributable to SSC dysfunction as opposed to dysfunction in more mature osteoblasts, and the overall contribution of SSCs as sources of osteoclastogenic signals is unclear.

Response to comment 16: We agree with the Reviewer that the data shows an increased osteoclastogenic potential of the aged bone marrow. Here, we meant to demonstrate that aged bone marrow contains increased osteoclast potential and activity which can be triggeredby the presence of Rankl and Csf1 alone. Together with the specific changes of Rankl and Csf1 expression in SSC lineage cells and the higher abundance of inflammatory stroma implies a connection between these cells and increased osteoclastogenesis. To clarify a direct link we have now conducted co-culture experiments as described in the preceding comment to show that aged SSC derived stroma drives higher monocytic output (Fig. 2k-n, see also response to comment 15).

17. Similarly, it is well established that Csf1 is a key osteoclastogenic factor and that the outcomes seen in Fig 5e-h all reflect the known and “generic” roles of Csf1 in osteoclastogenesis in these settings. These experiments do not specifically implicate CSF1 as a mediator of aging associated SSC dysfunction. Similar concerns apply to Fig 6a, as these results reflect the known general effects of BMP2 and CSF1 blockade and not a specific interaction indicating aging associated stem cell dysfunction.

Response to comment 17: We agree that the mechanism of actions of the two targeted factors are generally known. However, the novelty of our findings pertains to the specific roleof defined SSC lineage cells and the fact that we find that it is necessary to temper the increased prevalence of stroma-derived Csf1 with age and to simultaneously re-activate SSCs by addition of BMP2 in order to re-establish youthful fracture healing, which has previously not been described. In our latest addition of 10X single cell RNAseq data we show that combinatorial treatment with aCsf1/Bmp2 reduces osteoclast maturation in the hematopoietic compartment and increases skeletogenic gene expression in the mesenchymal compartment (Fig. 4)



18. The claim in the title is not clearly supported by the data presented. It is clear that aged skeletal stem cells display dysfunction in terms of osteogenic output, however, no direct experimental evidence is offered showing that an inflammatory niche is causal for these defects. Similarly no functional data is presented showing that SSCs are themselves the source of inflammatory factors crucial to aging associated skeletal pathologies.

Response to comment 18: We show that SSCs and their derived stroma express a numberof pro-inflammatory genes that are specifically upregulated with age (Fig. 3, Extended Data Fig. 5,6). Our co-culture experiments further provide functional evidence that these skeletal lineage cells drive myeloid expansion. The elevated prevalence of myeloid cells in turn mightfurther drive systemic inflammation as seen in our serum analysis of young and aged mice (Extended Data Fig. 6e-f). Indeed, there may be other cells and factors involved, which we now have acknowledged in the discussion (lines 410-413 and 435-437), but we also show that targeting SSCs by Bmp2 and SSC-lineage derived factors by anti-Csf1 is sufficient for rescue of age impaired fracture regeneration, promoting the fact of an essential role for SSCs and their signaling in regulating inflammatory osteoclastogenesis.

Minor points:

Minor 1. Page 5, line 99, the language used implies that the prior published papers on human and mice identify human and murine versions of the same cell type. However, given the substantial differences in surface immunophenotype between these cell types and also possible functional differences, it is unclear if these two reports are converge on a single cell type with species specific differences or distinct cell types in each study. Commentary and an update of the language used here would be critical in avoiding confusion about the currentstate of evidence regarding these two important studies. This issue may also have implications for the human relevance of this study that will be important to acknowledge.

Response to minor comment 1: We would like to clarify that the approach to identify human SSCs also relied on single cell whole transcriptome comparisons with purified mouse SSCs which has been extensively described in our paper (Chan et al 2018, particularly Figure 1, PMID: 30241615). We then used spatial and functional comparisons to confirm that we identified a homologous lineage tree in mice and men. We have also explored key differencesin mouse and human SSCs that might contribute to differences in skeletal growth and size. We have also recently used mouse and human SSCs to describe injury responses and articular cartilage regeneration (Murphy et al 2020, PMID: 32807933, Ambrosi et al 2020, PMID:32537886). Both identified cell types behave the same upon stimulation with the same factors, express the same master regulators such as Osterix, Runx2 and Sox9, as well as receptors to Bmp2, Ihh, Pth1r and are functionally homologous with each other (Chan et al 2018, PMID: 30241615).

Minor 2. The studies in the rainbow mice are helpful supporting data, but clonal clusters as suggested by this method only indicate clonal proliferation, which is not necessarily a stem cell property. This result could reflect proliferation of any later non-stem progenitor in the differentiation hierarchy. Updating the language on page 6 line 107 ("stem cell activity") is needed to reflect this.

Response to minor comment 2: We agree that clonal proliferation is not necessarily a stem cell property by itself, however, it is one of the hallmark stem cell features. Although Rainbow clones do not necessarily reflect SSC populations, we used it in combination with FACS analysis to denote the presence of SSC mediated clonal proliferation in the fracture calluses. Accordingly, we have updated the language on page 6 lines 112-114 (“Previously, we demonstrated location of clonal skeletogenesis, consistent with the detection of SSCs by FACS and the presence of stem cell activity, in the growth plates of postnatal day 3 long bones using Actin-CreERT Rainbow (“Rainbow”) mice, …”) (Chan et al 2015 PMID:25594184; Marecic et al 2015 PMID: 26216955)

Minor 3. Clarification of whether a true loss of the growth plate "page 5, line 82-3" versus an attenuation is observed at the ages examined would be helpful for placing results in context. Typically murine systems are considered as displaying a progressive attenuation but not a full loss of the growth plate with aging, such as in Mizuhashi et al. Nature 2018.

Response to minor comment 3: We agree that the remnants of a growth plate might still be observable which however reflects a virtual loss of it (Extended Data Fig. 1a,d). We have changed the phrasing in the updated manuscript to attenuation (line 86). Further, Mizuhashi et al. Nature 2018 only looked at 1-year old mice, which we found are not always reflective of an aged phenotype (data not shown), where the growth plate is still much more visible than in 24-mo mice.

Minor 4. It is difficult to interpret extended fig 2d given that the high numbers in the Thy1+ fracture samples compress the axis for all other samples. Separate graphs or a broken axis are needed.

Response to minor comment 4: The figure has been updated as part of the updated manuscript.

Minor 5. Why were NSG hosts used for study of murine SSCs? Typically that profound degreeof immunodeficiency isn't necessary for this study design. This is a relevant issue as inflammatory mechanisms of SSC dysfunction are highlighted as a mechanism and may be impacted in NSG hosts.

Response to minor comment 5: We have used NSG mice to minimize extrinsic factors such as immune responses to GFP in this assay. Since our scientific question is related to intrinsic alterations of SSC we believe this is the most faithful way to analyze it.

6. The emphasis in interpretation in the Fig 3 data is that young blood does not rescue aged SSC cellular defects. However, the most striking effect seen overall is that, conversely, old blood appears to be sufficient to confer cellular dysfunction into young SSCs (such as comparing IY to HY in Fig 3b-d). This undercuts the major conclusion of this section of the manuscript, that "Instead, cell intrinsic changes to SSC activity seem to be the primary driver in the age-related functional decline of the skeleton", as cell extrinsic factors indeed appear to be sufficient to drive SSC dysfunction.

Response to minor comment 6: It is entirely possible that a shift in circulating factors over time establishes an intrinsically changed state in aged SSCs. Young SSCs seem to be easily susceptible to an altered microenvironment (and stress as seen with scRNAseq data), such as during heterochronic parabiosis. Future studies will have to elucidate how that intrinsically altered state is established that on the other hand seems to be reversible by a local supraphysiological load of specific factors, i.e., anti-Csf1 and rBMP2.

7. The emphasis in the text describing Fig 3 on whether or not HA parabionts are fully restoredto the point of being similar to IY controls is puzzling, as HA parabionts could display substantial rescue without being comparable to IY controls. Comparison to IA controls would seem more relevant.

Response to minor comment 7: Statistical testing between all groups was conducted, but only significant differences between IY and heterochronic groups were found. The figures and legends have been updated to clarify this.

Minor 8. It is unclear how the conclusion that "SSC aging might therefore be the consequenceof a priming event that institutes an intrinsically aged state which renders these cells henceforth insensitive to blood-borne factors," follows from the immediately preceding statement that circulating factors exert a negative effect on the SSC lineage. Conversely, it would seem that the most straightforward interpretation of the Fig 3 data would be that circulating cell extrinsic factors associated with aging are indeed sufficient to induce SSC dysfunction, but that circulating factors from young mice are not sufficient to reverse dysfunction in aged SSCs. The claim that aging is intrinsic to SSCs appears to be at odds with the SSC dysfunction observed in HY mice relative to IY controls in Fig 3b-d. This wouldimply a mixed model with both cell intrinsic and cell extrinsic components to the SSC aging phenotype. Another possibility is that while there are circulating factors that confer aspects of "oldness" to SSCs, there may simply not be corresponding circulating factors that can confer "youngness".

Response to minor comment 8: The fact that young SSCs are affected by factors of old blood does not exclude the fact that aged cells are not susceptible to blood borne factors. As stated above an intrinsically altered state might be acquired over time that could have been initiated by extrinsic changes. In the artificial setting of parabiosis, the reviewer is correct, there seems to bea dominant effect of aged factors which can be seen as an effect on the young SSC lineage but not vice versa as an effect of young blood on aged SSCs. The possibility that "youngness" of SSCs cannot be conferred by circulating factors as easy as "oldness" to SSCs further supports that assumption. Follow up studies will have to address the specific mechanisms driving aging ofSSCs.

Minor 9. Regarding the description of cluster associated genes in page 10 line 209-11, Jundis a suppressor of osteogenesis as in Kawamata et al. J Cell Biochem, which would appear to run counter to the description in the text.

Response to minor comment 9: We have now updated our single cell RNAseq analysis to better reflect differences in subsets of SSCs. As pointed out by reviewer 1 and supported by literature, expressions of genes like *Jund* are most likely stress-related stemming from cell processing prior to sequencing, since other stress-related genes are found to be expressed in the same cells (Denisenke et al 2020, PMID:32487174; van den Brink et al 2017, PMID: 28960196).

Our updated analysis now better reflects cluster specific genes and infers functional differences (Fig. 3 and Extended Data Fig. 5, see also response to comment 9).

Minor 10. BMP2 plays complex roles in both anabolic and catabolic responses, so simply describing it as a gene associated with bone formation as on page 11 line 217 is potentially misleading. Similarly, the anabolic versus catabolic effects of TGF-beta on bone metabolism depend on amount and context (such as in Erlebacher et al, J Cell Biol 1996) in contrast with the statement on page 12 line 244.

Response to minor comment 10: We agree that there are context specific differences of these genes, however, in our past published work we have clearly shown that Bmp2 induces osteogenic output in vitro and in vivo (Chan et al. 2015, 2018 and Murphy et al 2020). We have revised this paragraph with our updated analysis and removed Bmp2 to avoid controversy.

Minor 11. Relevant to the statement that “Rankl by bone marrow stromal cells leads to enhanced osteoclastogenesis and adipogenesis at the expense of osteogenesis" (p16, line 344-5), while RANKL levels may increase with aging and relative amounts of adipogenesis increase with aging, this reviewer is not aware of any clear or direct evidence that increases RANKL expression specifically drive the increases in adipogenesis at the expense of osteogenesis observed with aging.

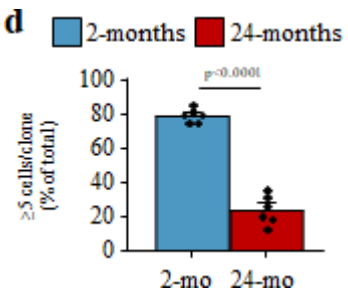
Response to minor comment 11: It is generally accepted that marrow adipogenesis and osteogenesis are inversely correlated. The statement was mainly based on the references provided (PMID: 24753250 as well as PMID: 28162969). We now have rephrased to state: “Increased expression of Rankl by bone marrow stromal cells has been shown to be a sign of elevated bone marrow adipogenesis which in turn could promote osteoclastogenesis” (lines 416-418).

Minor 12. Extended Data 2d and 2e are not mentioned in the main text.

Response to minor comment 12: All figures have been checked to be mentioned in the text.

Minor 13. In Figure 1d, how was the calculation for the number of clones made? The standard deviation is per group is missing in the plot.

Response to minor comment 13: The analysis has been re-done to now show differencesin number of clones with cells equal or larger than 5 versus number of clones consisting of less than 5 cells (Fig. 1d). Six distinct callus regions (5-19 clones per section) from two miceper age group were counted. Statistical analysis has been conducted as described in figure legend.



Minor 14. In the main text Line 123, the figure is mislabeled as 1d-e, instead it should be 2d-e.

Response to minor comment 14: Figure references have been updated with new Figures.

Minor 15. The TRAP staining image in Fig 2d is unclear. A better image, including a lower power view is needed.

Response to minor comment 15: Image has been updated.

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have adequately addressed my concerns. I support publication of this manuscript in its current form.

Referee #2 (Remarks to the Author):

The authors addressed extensively my comments, and also, in my view, responded adequately to comments from other reviewers. The experiments described in the paper are very useful and important, the interpretation is interesting and open many opportunities for follow up. This is a very strong candidate for publication - more data should not be required and remaining open questions should be addressed in future work.

Referee #3 (Remarks to the Author):

In this revised version of the Ambrosi et al. manuscript, some of the concerns raised have been adequately addressed. Some concerns remain, though some of these remaining concerns can be potentially addressed with text changes and qualification of conclusions. Detailed comments are provided below.

1. As acknowledged by the authors in the response to comment 4, it remains the case that it is unclear to what degree the aging associated defects truly occur specifically at the level of SSCs versus being at the level of SSC-derived, more mature populations. This is emphasized by the use of the term "SSC/BCSP" in the response given. While SSC frequency defects presumably must occur at the level of

the SSC itself, other defects, such as reduction of ossicle volume could be due to defects in more mature, non-SSC populations. Discrimination of defects at the stem cell versus more mature level is important to justify the novelty of the work by discriminating the present study from the extensive literature on aging associated defects in the osteoblast compartment. In vitro differentiation assays as in Fig 1l can reflect defects in later post-SSC populations. Similarly, in vitro differentiation has a poor correlation with in vivo lineage allocation and thus these studies are at best secondary supporting data for defects in SSCs. CFU-F assays similarly have an unclear relationship to stemness, as in most CFU-F assays it cannot be distinguished if the cells actually proliferating still retain a stem phenotype or if they have matured during the CFU-F assay culture process. Similar issues apply to the use of confetti based analysis, where the proliferation driving the cluster formation could largely reflect proliferation in a non-stem population similar to transit amplifying cells described in other systems. Competitive self-renewal assays would be feasible and would not require use of single clones and would indicate a defect that can only be ascribed to SSCs and not more mature populations. Instead, along the lines of prior comment #5, co-transplantation of differentially labeled young and old SSCs could be used in an ossicle system to assess self renewal using the methods previously published by the authors. These young and old SSCs could be initially transplanted at a 1:1 ratio and a progressive drift in this ratio in favor of the young SSCs would provide compelling evidence of a self-renewal defect in old SSCs.
Ultimately, it is appreciated that it is challenging to distinguish defects specifically in the stem cell from defects in the function of downstream stem cell-derived populations and that reasonable attempts have been made to address this issue in the initial submission. If a competitive renewal or similar assay is not performed, there should at least be clear qualification in the text that most or all of the assays performed here to demonstrate aging associated SSC defects cannot distinguish between defects specific to SSCs versus functional defects in more mature, non-stem SSC derived populations.

2. Prior comment 11, the requested dynamic histomorphometry was not performed, and this is a standard approach needed to discriminate the anabolic versus catabolic contributions of a phenotype. Analysis of circulating CTX levels in heterochronic aged versus isochronic aged mice is problematic as presumably the CTX levels in heterochronic pairs reflect an average of the two parabionts, and thus parabiosis induced changes in the aged parabiont will be conflated with the effect of averaging the CTX levels in aged versus young parabionts. Ultimately, however this weakness is partially offset by the bone spicule formation assays where the total endpoint bone volume can be inferred to be largely reflective of de novo bone formation. If histomorphometry is not performed, it must be acknowledged that the relative contribution of anabolic versus catabolic mechanisms to dysfunction by aged SSCs has not be resolved.

3. Comment 13: " In our experience trabecular bone mass is a superior predictor of bone brittleness in mice." This claim is not in line with longstanding and accepted literature in the field. It is well-accepted that cortical bone is overall the major contributor to the biomechanical resistance to fracture and an overall more important role in determining the biomechanical properties of bone than trabecular bone (e.g., Holzer et al. JBMR 2009, Ito et al. Bone 2002 and others). It is also not the case that both BMD and BV/TV are accepted reporting parameters as stated. BMD and BV/TV provide very different information and are not interchangeable, and the BMD values derived here by uCT are very different in interpretation from the values obtained by DXA, the latter of which are a more integrative measure of bone architecture. The interpretation of BMD is also arguably less straightforward than BV/TV as BMD depends on factors such as the time since bone matrix secretion and rate of turnover as much as the overall anabolic activity. Accurate reporting of BMD requires regular calibration of the instrument used specifically for this purpose using phantom targets with graded densities. What calibration procedures were performed? Ultimately, it is recommended that BMD values reported be moved to supplemental figures and that phenotypic reporting focus on the more standard architectural parameters of trabecular BV/TV and cortical thickness. Bouxsein et al. JBMR 2010 provides a summary of preferred reporting and methods documentation for uCT. When reporting fracture phenotypes, the callus should be contoured as a region of interest conforming to the callus volume that avoids the underlying cortical bone and callus total BV should be reported as the major endpoint.

4. Based on the response to minor point 1, it appears that while human and murine SSCs share some properties, it is still uncertain whether these are fundamentally the same cell type. It is therefore

recommended that this be made clear in the relevant portions of the text to avoid confusion in the field on this important point, as this could lead to unnecessarily conflating these two potentially distinct cell types.

5. Minor comment 5: Use of hosts tolerized to GFP via expression of a GFP variant in a non-relevant location would be preferred over NSG hosts as a method to exclude immune responses directed against graft GFP. However, it is acknowledged that this is unlikely to fundamentally change experimental results and avoiding anti-graft GFP responses is important. Use of strongly immunosuppressed hosts should be acknowledged as a limitation where relevant, especially considering that inflammatory mechanisms are invoked as driving aging associated SSC dysfunction.

6. As stated in prior minor comment 7, the relevant comparison for arguing that aged SSCs are not influenced by extrinsic factors is comparison of HA to IA mice, not HA to IY mice as emphasized in the text.

7. Prior Comment 10: Concerns about dropping CD200 from the SSC definition raised by two reviewers and not strongly addressed during revisions. Claims that all cells within the Lin-6C3-Thy-CD105- compartment are CD200+ outside of neonatal or fetal contexts is contradicted by other publications, such as Han et al. JCI 2019, Matthews eLife 2021 and others. Are the pre-BCSPs now unable to be discriminated from SSCs and conflated as one population? The single cell RNA-seq studies performed do suggest substantial heterogeneity in the SSC pool analyzed here and certainly raise the possibility that many of the cells within the SSC definition here are not stem cells. Resolving the identity of SSCs versus other skeletal populations and how aging may change the definition of an SSC or drive other non-stem populations to assume the SSC immunophenotype, if not the actual functional characteristics of SSCs, is central to the claims of this manuscript. Definition and identification of SSCs are prerequisites to map aging associated physiologic defects back to this specific cell type. However, it is appreciated that a comprehensive treatment of this topic would perhaps require another manuscript's worth of data. Thus, a focused consideration of this issue would be sufficient. It may be useful to use an isotype matched control with the same fluorophore as CD200 given the importance of this issue both for the present manuscript and for the broader field. The use of contour plots should be avoided for FACS data, as these can be misleading. Dot plots are preferred. It should also be at least entertained whether preBCSPs and SSCs can be here can be discriminated based on the degree of CD200 expression, as CD200 bright versus CD200 dim populations.

8. As acknowledged by the language used in the manuscript text, altered fracture responses in CSF1 +/- bones appear to reflect the well known osteopetrosis phenotype seen with CSF1 deficiency and the well known role of osteoclasts in fracture healing. These relationships between CSF1 and osteoclast development and osteoclast deficiency/osteopetrosis and impaired fracture healing are all well described and are not aging dependent.
The presence of this osteopetrosis phenotype prevents this model from being informative about mechanisms of aging. For the same reason, CSF1 +/- mice display increased bone mass due to their defects in osteoclastogenesis regardless of the role of CSF1 in aging associated pathologies. This issue also explains why high level CSF1 antagonism did not rescue fracture healing and CSF1 deficient mice display impaired biomechanical outcomes of fracture healing--as the resulting blockade in osteoclast generation produces a known defect in fracture responses that is independent of aging associated factors. Thus, most of the in vivo data offered regarding CSF1 appears to reflects the known and generic role of osteoclasts in the systems examined, except perhaps the results of the low level CSF1 antagonism group in Fig 4.

Author Rebuttals to Second Revision:

Referee #1 (Remarks to the Author):

The authors have adequately addressed my concerns. I support publication of this manuscript in its current form.

Response to Reviewer #1: We would like to thank the reviewer again for thoughtful comments and suggestions helping us to improve our manuscript.

Referee #2 (Remarks to the Author):

The authors addressed extensively my comments, and also, in my view, responded adequately to comments from other reviewers. The experiments described in the paper are very useful and important, the interpretation is interesting and open many opportunities for follow up. This is a very strong candidate for publication - more data should not be required and remaining open questions should be addressed in future work.

Response to Reviewer #2: We appreciate the reviewer's support and enthusiasm for our work.

Referee #3 (Remarks to the Author):

In this revised version of the Ambrosi et al. manuscript, some of the concerns raised have been adequately addressed. Some concerns remain, though some of these remaining concerns can be potentially addressed with text changes and qualification of conclusions. Detailed comments are provided below.

Response to Reviewer #3: We thank the reviewer for the initial comments and suggestions and greatly appreciate the time taken for following up on remaining concerns of the revised manuscript. We have carefully read the comments and addressed each in our responses. We hope the reviewer will now deem this work acceptable for publication.

1. As acknowledged by the authors in the response to comment 4, it remains the case that it is unclear to what degree the aging associated defects truly occur specifically at the level of SSCs versus being at the level of SSC-derived, more mature populations. This is emphasized by the use of the term "SSC/BCSP" in the response given. While SSC frequency defects presumably must occur at the level of the SSC itself, other defects, such as reduction of ossicle volume could be due to defects in more mature, non-SSC populations. Discrimination of defects at the stem cell versus more mature level is important to justify the novelty of the work by discriminating the present study from the extensive literature on aging associated defects in the osteoblast compartment. In vitro differentiation assays as in Fig 1l can reflect

defects in later post-SSC populations. Similarly, in vitro differentiation has a poor correlation with in vivo lineage allocation and thus these studies are at best secondary supporting data for defects in SSCs. CFU-F assays similarly have an unclear relationship to stemness, as in most CFU-F assays it cannot be distinguished if the cells actually proliferating still retain a stem phenotype or if they have matured during the CFU-F assay culture process. Similar issues apply to the use of confetti based analysis, where the proliferation driving the cluster formation could largely reflect proliferation in a non-stem population similar to transit amplifying cells described in other systems. Competitive self-renewal assays would be feasible and would not require use of single clones and would indicate a defect that can onlybe ascribed to SSCs and not more mature populations. Instead, along the lines of prior comment #5, co-transplantation of differentially labeled young and old SSCs could be used in an ossicle system to assess self renewal using the methods previously published by the authors. These young and old SSCs could be initially transplanted at a 1:1 ratio and a progressive drift in this ratio in favor of the young SSCs would provide compelling evidence of a self-renewal defect in old SSCs.
Ultimately, it is appreciated that it is challenging to distinguish defects specifically in the stem cell from defects in the function of downstream stem cell-derived populations and that reasonable attempts have been made to address this issue in the initial submission. If a competitive renewal or similar assay is not performed, there should at least be clear qualification in the text that most or all of the assays performed here to demonstrate aging associated SSC defects cannot distinguish between defects specific to SSCs versus functional defects in more mature, non-stem SSC derived populations.

Response to Comment #1: We thank the reviewer for this comment. We would like to re-iterate that we observe significant changes to the frequency, activation state, and proliferative capacity at the level of the stem cell (SSC) and its downstream multipotent progenitor (BCSP) (Figure 1). In both the stem cell and its downstream progenitor, we see similar defects. The inheritence of some age-related molecular changes we observe between SSC and BCSP, that is, a stem cell and its derivative cell is also noted in a recent study in the hematopietic system in which age-related changes in activity was observed in the HSC and its derivative MPP (PMID: 27811054). This is an important observation as it suggests that defects at the stem cell level could be downstream, perhaps to the level of mature lineages. As described by the reviewer we conducted all commonly used assays to interrogate skeletal stem cell properties. We agree that all assays come with certain caveats but as a whole clearly suggestthat the initial age-related defects described originate at the apex of the lineage tree. We acknowledged in our response to the reviewer that defects likely are inherited (and maybe even accumulated) during downstream differentiation (see also responses to comment #1, 4,5 of first revision). Future studies beyond the scope of this manuscript will hopefully be able to resolve this issue definitively. We have now added a paragraph to the discussion highlighting the concern of the reviewer in response to comment 1 and 2: “While our parabiosis studies could not entirely resolve relative contributions of anabolic versus catabolic mechanisms leading to skeletal dysfunction by aged SSCs, future studies will also have to

assess the degree of bone-forming defects on the level of downstream cell populations that might inherit or even accumulate age-related impairments." (lines 406-409).

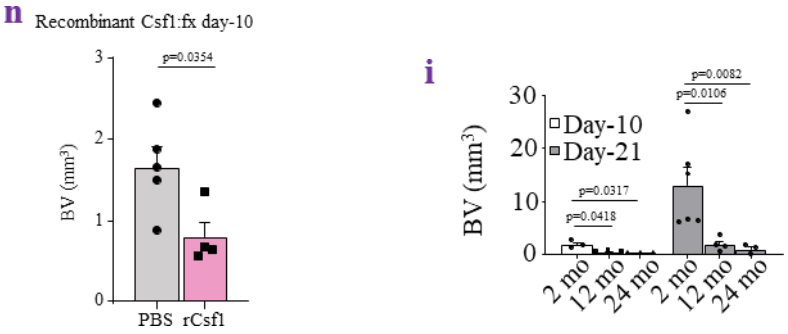
2. Prior comment 11, the requested dynamic histomorphometry was not performed, and this is a standard approach needed to discriminate the anabolic versus catabolic contributions of a phenotype. Analysis of circulating CTX levels in heterochronic aged versus isochronic aged mice is problematic as presumably the CTX levels in heterochronic pairs reflect an average of the two parabionts, and thus parabiosis induced changes in the aged parabiont will be conflated with the effect of averaging the CTX levels in aged versus young parabionts. Ultimately, however this weakness is partially offset by the bone spicule formation assays where the total endpoint bone volume can be inferred to be largely reflective of de novo bone formation. If histomorphometry is not performed, it must be acknowledged that the relative contribution of anabolic versus catabolic mechanisms to dysfunction by aged SSCs has not be resolved.

Response to Comment #2: We would like to apologize for not performing some of the additional requested analysis which would require new parabiotic experiments. Due to reasons related to animal welfare, cost, and access restrictions during Covid-19 regulations, we were not able to perform these additional parabiosis experiments that would have been able to address this issue in a satisfying manner. As suggested by the reviewer we have added a sentence in the discussion to acknowledge that concern: "While our parabiosis studies could not entirely resolve relative contributions of anabolic versus catabolic mechanisms leading to skeletal dysfunction by aged SSCs, future studies will also have to assess the degree of bone-forming defects on the level of downstream cell populations that might inherit or even accumulate age-related impairments." (Lines 406-409).

3. Comment 13: " In our experience trabecular bone mass is a superior predictor of bone brittleness in mice." This claim is not in line with longstanding and accepted literature in the field. It is well-accepted that cortical bone is overall the major contributor to the biomechanical resistance to fracture and an overall more important role in determining the biomechanical properties of bone than trabecular bone (e.g., Holzer et al. JBMR 2009, Ito et al. Bone 2002 and others). It is also not the case that both BMD and BV/TV are accepted reporting parameters as stated. BMD and BV/TV provide very different information and are not interchangeable, and the BMD values derived here by uCT are very different in interpretation from the values obtained by DXA, the latter of which are a more integrative measure of bone architecture. The interpretation of BMD is also arguably less straightforward than BV/TV as BMD depends on factors such as the time since bone matrix secretion and rate of turnover as much as the overall anabolic activity. Accurate reporting of BMD requires regular calibration of the instrument used specifically for this purpose using phantom targets with graded densities. What calibration procedures were performed? Ultimately, it is

recommended that BMD values reported be moved to supplemental figures and that phenotypic reporting focus on the more standard architectural parameters of trabecular BV/TV and cortical thickness. Bouxsein et al. JBMR 2010 provides a summary of preferred reporting and methods documentation for uCT. When reporting fracture phenotypes, the callus should be contoured as a region of interest conforming to the callus volume that avoidsthe underlying cortical bone and callus total BV should be reported as the major endpoint.

Response to Comment #3: We apologize if we have not clarified in enough detail. All measurements on BMD were performed based on phantom calibration (lines 774-775). Generally, CT measurements were performed in reference to guidelines by Bouxsein et al. JBMR 2010. As mentioned by the reviewer, here we assessed fracture callus phenotypes by contouring region of interest as the callus area outside the boundaries of cortical bone. We have amended the method section accordingly (Lines 780-781). We have also updated all fracture callus measurements for BV/TV to now display BV, which has not changed the results of those experiments (Fig. 3n and Extended Data Fig. 1i).



4. Based on the response to minor point 1, it appears that while human and murine SSCs share some properties, it is still uncertain whether these are fundamentally the same cell type. It is therefore recommended that this be made clear in the relevant portions of the text to avoid confusion in the field on this important point, as this could lead to unnecessarily conflating these two potentially distinct cell types.

Response to Comment #4: We kindly disagree. We have elaborated on why we think the two populations are the same cell type, as it was extensively explored and confirmed in our studies (Chan et al 2018, Cell and Murphy et al 2020, Nature Medicine, Ambrosi et al 2020, Aging Cell). However, we note that mice and human SSCs *are* different in some regulatory aspects (see Chan et al 2018) and therefore mechanims related to aging mouse SSCs will need to be independenly validated in human SSCs. We now include the following statementto express this caveat: “Our findings showcase the multifaceted nature of SSC aging where no single causative factor is responsible for the functional decline of these cells. Previous work also established a degree of divergence in respect to regulatory mechanisms betweenmice and human SSCs[13]. Thus, we anticipate that a comprehensive comparative study

between equivalent populations of SSC lineages in mice and humans, two vertebrates with distinct growth and aging rates, could elucidate the full spectrum of deterministic mechanisms that underlie skeletal aging." (Lines 445-450).

5. Minor comment 5: Use of hosts tolerized to GFP via expression of a GFP variant in a non-relevant location would be preferred over NSG hosts as a method to exclude immune responses directed against graft GFP. However, it is acknowledged that this is unlikely to fundamentally change experimental results and avoiding anti-graft GFP responses is important. Use of strongly immunosuppressed hosts should be acknowledged as a limitation where relevant, especially considering that inflammatory mechanisms are invoked as driving aging associated SSC dysfunction.

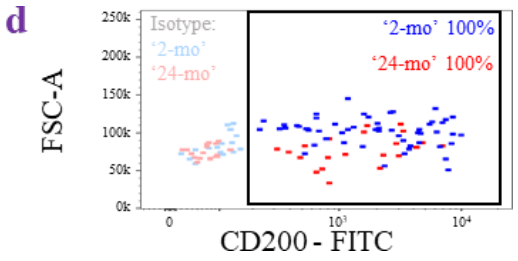
Response to Comment #5: Renal capsule transplants in NSG hosts mainly aimed to address the intrinsic capacity of differently aged SSCs to form ossicles. Tolerizing mice tendto yield variable results as readout depends on positive and negative selection of the developing immune repertoire against many different antigens. As the main immune reservoirof NSG mice constitutes neutrophils and monocytes, we were able to make the observation that SSCs from aged mice formed ossicles with less bone and increased osteoclast activity (Fig 1). We acknowledge that the absence of a complete immune system might have affectedthe outcome of the results related to bone resorption, but our conclusions were further validated with additional experiments. We have amended the following sentence which addresses this concern: "In summary, transplantation of aged SSCs and BCSPs into young mice did not improve their skeletogenic potential, indicating that cell-autonomous changes underlie their functional decline and suggesting that aged skeletal lineages produce factors that enhance bone resorption, at least in the setting of immunocompromised mice enriched for monocytic cells." (lines 157-160)

6. As stated in prior minor comment 7, the relevant comparison for arguing that aged SSCs are not influenced by extrinsic factors is comparison of HA to IA mice, not HA to IY mice as emphasized in the text.

Response to Comment #6: It is unclear to what the reviewer exactly is referring to. The textin lines 171-194 compares the relationship of HA to IA and IY to a similar extent. In regard to SSCs it for example reads: "SSC/BCSP frequencies also did not significantly improve in healing femora of HA parabionts compared to isochronic controls (Extended Data Fig. 4e). In addition, the in vitro osteogenic and chondrogenic potentials of SSCs isolated from HA calluses remained similar to IA derived SSCs and significantly decreased compared to IY SSCs (Fig. 2e-f)."

7. Prior Comment 10: Concerns about dropping CD200 from the SSC definition raised by two reviewers and not strongly addressed during revisions. Claims that all cells within the Lin-6C3-Thy-CD105- compartment are CD200+ outside of neonatal or fetal contexts is contradicted by other publications, such as Han et al. JCI 2019, Matthews eLife 2021 and others. Are the pre-BCSPs now unable to be discriminated from SSCs and conflated as one population? The single cell RNA-seq studies performed do suggest substantial heterogeneity in the SSC pool analyzed here and certainly raise the possibility that many of the cells within the SSC definition here are not stem cells. Resolving the identity of SSCs versus other skeletal populations and how aging may change the definition of an SSC or drive other non-stem populations to assume the SSC immunophenotype, if not the actual functional characteristics of SSCs, is central to the claims of this manuscript. Definition and identification of SSCs are prerequisites to map aging associated physiologic defects back to this specific cell type. However, it is appreciated that a comprehensive treatment of this topic would perhaps require another manuscript's worth of data. Thus, a focused consideration of this issue would be sufficient. It may be useful to use an isotype matched control with the same fluorophore as CD200 given the importance of this issue both for the present manuscript and for the broader field. The use of contour plots should be avoided for FACS data, as these can be misleading. Dot plots are preferred. It should also be at least entertained whether preBCSPs and SSCs can be here can be discriminated based on the degree of CD200 expression, as CD200 bright versus CD200 dim populations.

Response to Comment #7: We have replaced the contour plot in Extended Data 2d with a dotplot which now also shows isotype controls performed on cells from both ages instead of FMO. We have also amended the discussion speculating about varying CD200 levels and their putative role in differentiating SSC and pre-BCSP populations. “Of note, in contrast to our initial report phenotypic SSCs from adult mice could not be discriminated from pre-BCSPby CD200 positivity but might rather be distinguishable by the degree of CD200 expression[2].”(lines 380-382).



8. As acknowledged by the language used in the manuscript text, altered fracture responses in CSF1 +/- bones appear to reflect the well known osteopetrosis phenotype seen with CSF1 deficiency and the well-known role of osteoclasts in fracture healing. These relationships between CSF1 and osteoclast development and osteoclast deficiency/osteopetrosis and impaired fracture healing are all well described and are not aging dependent.

The presence of this osteopetrosis phenotype prevents this model from being informative about mechanisms of aging. For the same reason, CSF1 +/- mice display increased bone mass due to their defects in osteoclastogenesis regardless of the role of CSF1 in aging associated pathologies. This issue also explains why high level CSF1 antagonism did not rescue fracture healing and CSF1 deficient mice display impaired biomechanical outcomes of fracture healing--as the resulting blockade in osteoclast generation produces a known defect in fracture responses that is independent of aging associated factors. Thus, most of the in vivo data offered regarding CSF1 appears to reflect the known and generic role of osteoclasts in the systems examined, except perhaps the results of the low level CSF1 antagonism group in Fig 4.

Response to Comment #8: We thank the reviewer for this comment. In order to arrive at the final and important conclusion that low level CSF1 is beneficial in the combined treatment wefeel it is important to show where our independently drawn conclusions have originated from.Even though some of those results might be anticipated from previous studies on osteoclastsand CSF1 that we cited, we note that many inflammatory cytokines display differential expression between young and aged SSC-derived stromal cells (see Fig. 3e). While it was not obvious to us from the initial gene expression data which of these multiple candidates may be suitable to ablate, the $CSF1^{+/-}$ studies indicated that supressing CSF1 levels might give a noticeable in vivo result. It is also important to note that CSF1 itself was not sufficient to restore bone healing, instead it required the combined effect of BMP2 and repressing highCSF1 levels. We are not aware of studies that address this issue in a combined approach similar to ours.

Reviewer Reports on the Third Revision:

Referee #3 (Remarks to the Author):

Prior comments have been adequately addressed with only one remaining issue. While it is appreciated in prior comment 8 that the in vivo studies on CSF1-deficient mice may have been helpful for prioritization of candidate mediators, it remains the case that the results of these studies in germline CSF1 haploinsufficient mice likely reflect the general role of CSF1 in osteoclast generation and not any specific function of CSF1 related to aging. At least no data currently offered specifically with respect to these studies in CSF1 null allele mice allow for clear discrimination of a specific role of CSF1 in aging versus the known and general role of CSF1 in osteoclastogenesis. How this issue impacts the interpretation of studies in CSF1 haploinsufficient mice should be acknowledged or addressed in some fashion in the presentation of the data.

Author Rebuttals to Third Revision:

Referee #3 (Remarks to the Author):
Prior comments have been adequately addressed with only one remaining issue. While it is appreciated in prior comment 8 that the in vivo studies on CSF1-deficient mice may have been helpful for prioritization of candidate mediators, it remains the case that the results of these studies in germline CSF1 haploinsufficient mice likely reflect the general role of CSF1 in osteoclast generation and not any specific function of CSF1 related to

aging. At least no data currently offered specifically with respect to these studies in CSF1 null allele mice allowfor clear discrimination of a specific role of CSF1 in aging versus the known and general roleof CSF1 in osteoclastogenesis. How this issue impacts the interpretation of studies in CSF1 haploinsufficient mice should be acknowledged or addressed in some fashion in the presentation of the data.

Response to Reviewer #3: We would like to thank the reviewer again for helpful suggestionsand address the final comment. In order to acknowledge the concern we have included the interpretation that the Csf-mouse phenotype might not be age-specific in lines 246-248 of thefinal manuscript: "Even though not necessarily age-related, we also found that 15-month-oldhaplo-insufficient *Csf1* mice had superior bone parameters, independent of sex."