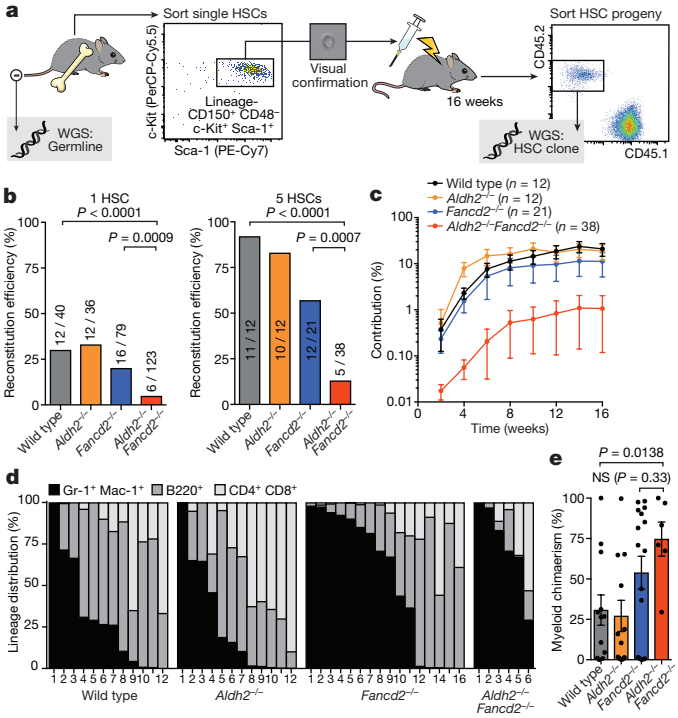

Figure 4 | Single HSC transplantation reveals that *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs are functionally compromised. a, Transplantation of single HSCs for the generation of HSC clones *in vivo*. The HSC progeny (CD45.2^{+}) were recovered after four months and analysed by whole-genome sequencing, alongside a germline reference. **b**, Percentage and number of irradiated recipients that were positive for reconstitution by one or five transplanted HSCs (*P* calculated by two-sided Fisher's exact test). **c**, Contribution to blood production over time by five transplanted HSCs (data shown as mean and s.e.m.). **d**, Lineage composition of single-HSC clones four months after transplant; columns represent individual recipients or clones. **e**, Proportion of myeloid (Gr-1^{+}Mac-1^{+}) HSC-derived white blood cells (*P* calculated by two-sided Mann–Whitney test; data shown as mean and s.e.m.; $n = 12$, 12, 16 and 6 HSC clones, left to right).

compartment. This is a critical question because there is evidence that HSCs differ in their DNA repair capacity and response compared to later progenitors[11]. Two obstacles had to be overcome in order to establish whether endogenous aldehydes mutate the genomes of these vital cells. First, the stochastic nature of DNA damage makes it unlikely that the same mutation will occur in multiple cells. Second, the scarcity of HSCs, especially in the case of *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice, precludes the use of most conventional techniques to assess DNA damage. We also wanted to ascertain whether mutations arise in functional stem cells, and therefore avoided whole-genome amplification or short-term *in vitro* expansion of cells isolated by flow cytometry. Instead, we decided to define HSCs functionally and exploit the ability of a single HSC to reconstitute long-term blood production following transplantation into a lethally irradiated mouse[20].

Our approach combines transplantation of single HSCs with whole-genome sequencing to obtain the mutational landscape of stem cells, while also allowing us to assess the functional capacity of mutant HSCs (Fig. 4a). We carried out transplants with one or five *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs (Fig. 4b). These stem cells rarely engrafted (with a frequency of 4.8%), contributed less to haematopoiesis and were myeloid-biased compared to controls (Fig. 4b–e). These results indicate that *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs are severely functionally compromised and share features with aged HSCs[21,22].

Mutational landscape of aldehyde-damaged stem cells

Our ultimate goal was to obtain clonal blood, which provided us with a physiological method to amplify stem-cell genomes. As outlined (Fig. 4a), four months after transplantation, we isolated the CD45.2^{+} HSC progeny and performed whole-genome sequencing at 20× coverage; tail DNA from the donor mouse served as the germline reference. This allowed us to detect heterozygous somatic changes, which are absent in the matched germline reference and represent mutations in the HSC. Genomes of *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs were mutated with increased prevalence of indels, rearrangements and translocations (Fig. 5a, Extended Data Fig. 4).

Although the number of single-base substitutions was significantly higher in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ genomes (Fig. 5b), the total numbers were low and no changes were detected in the type of substitutions (Fig. 5c). We also found no difference in the frequency or pattern of point mutations in bone marrow cells of *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice using the SelectcII Big Blue *in vivo* reporter assay (Extended Data Fig. 5).

A limitation of our approach is that cells with the capacity to engraft may represent the least mutated HSCs. Nevertheless, we observed significant increases in the frequency of deletions, which were more prevalent (Fig. 5d, e) and larger (Fig. 5f) in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ genomes. The mean variant allele frequency (VAF) for all filtered indels was 0.47, establishing that these changes are of clonal origin. By examining the flanking regions, we found that microhomology-mediated deletions are the main contributors to the mutations observed in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs, indicative of end-joining repair of DSBs[23] (Fig. 5g, h). Additionally, the increase in the size of the deletions (Fig. 5f) suggests a role for alternative end-joining in the repair of some of these breaks, as alternative end-joining is characterized by increased resection in comparison to classical NHEJ[24]. Next, we analysed the location of indels across the genome, as recent work has suggested a role for the Fanconi anaemia pathway in preventing genomic instability at transcription–replication collisions[25,26]. However, we found no evidence of microhomology-mediated deletions being enriched at coding regions or transcribed genes (Fig. 5i, j), suggesting that DSB formation in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs is stochastic.

The most striking change in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs was the presence of rearrangements that were not detected in most controls. *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ stem cells contained on average two rearrangements per genome; in contrast, we observed only two large deletions among all ten control HSC genomes (Fig. 5k–l). In summary, these data