

Early turbulent mixing as the origin of chemical homogeneity in open star clusters

Yi Feng¹ & Mark R. Krumholz¹

The abundances of elements in stars are critical clues to stars' origins. Observed star-to-star variations in logarithmic abundance within an open star cluster—a gravitationally bound ensemble of stars in the Galactic plane—are typically only about 0.01 to 0.05 over many elements^{1–9}, which is noticeably smaller than the variation of about 0.06 to 0.3 seen in the interstellar medium from which the stars form^{10–14}. It is unknown why star clusters are so homogenous, and whether homogeneity should also prevail in regions of lower star formation efficiency that do not produce bound clusters. Here we report simulations that trace the mixing of chemical elements as star-forming clouds assemble and collapse. We show that turbulent mixing during cloud assembly naturally produces a stellar abundance scatter at least five times smaller than that in the gas, which is sufficient to explain the observed chemical homogeneity of stars. Moreover, mixing occurs very early, so that regions with star formation efficiencies of about 10 per cent are nearly as well mixed as those with formation efficiencies of about 50 per cent. This implies that even regions that do not form bound clusters are likely to be well mixed, and improves the prospects of using 'chemical tagging' to reconstruct (via their unique chemical signatures, or tags) star clusters whose constituent stars have become unbound from one another and spread across the Galactic disk.

The question of how star clusters become chemically well mixed has received relatively little attention. With a few exceptions¹⁵, work to date has been limited to simple analytic estimates¹⁶, or to calculations omitting star formation and self-gravity^{17,18}. To improve this situation, we have performed a series of simulations of star cluster formation including hydrodynamics, gravity, and optically thin radiative heating and cooling. Our simulations use the ORION code^{19–21}, with a new implementation of particle-mesh gravity (Methods, Extended Data Figs 1 and 2). We use initial conditions based on the 'colliding flow' model²². We consider a region containing gas of number density $n_0 = 1 \text{ cm}^{-3}$ (mass density $\rho_0 = 2.1 \times 10^{-24} \text{ g cm}^{-3}$) with initial temperature $T = 5,000 \text{ K}$. The gas has a random turbulent velocity v_{rms} , for which we consider two values: 0.17 and 1.7 km s^{-1} ; we refer to runs with these values as S and L, for small and large turbulence, respectively. In addition to the turbulent velocity, we set up two cylindrical regions 32 pc long and 32 pc in radius, centred on the x axis, with their closer ends separated by 64 pc, within which the gas has a uniform velocity $v_0 = 9.2 \text{ km s}^{-1}$, directed towards the other cylinder. To trace chemical mixing, the simulation includes two passive scalars Q_L and Q_R , which have initial abundances of 1 within the left and right cylinders, respectively, and 0 elsewhere. In addition to these two simulations with smooth initial conditions, we also run a simulation C (for clumpy)¹⁵, in which we randomly add cold clumps one coarse cell in radius with a filling fraction of 0.05 and a number density $n_c = 132.5 \text{ cm}^{-3}$; at this density the equilibrium temperature is such that the pressure is in equilibrium with that of the warm low-density background. Full details of the simulations are given in Methods.

The overall evolution of our simulations is very similar to previous colliding flow simulations^{15,22,23} (Fig. 1). The two streams of gas converge rapidly, and compression of material at the heads of the two cylinders leads to thermal instability and the formation of a cold phase even before the two flows collide. For runs S and L, the two streams collide just before

10 Myr of evolution, and this produces a dense, cold, turbulent layer that is gravitationally unstable. The layer begins to form stars at ~ 19 Myr in run S, and approximately 50% of the gas in the two streams has been converted to stars by ~ 25 Myr of evolution (Extended Data Fig. 3). In run L, star formation begins at ~ 25 Myr, and follows a similar time evolution to run S thereafter. In run C, the two streams collide at ~ 6 Myr because the dense cold phase is less decelerated by the warm phase. Star formation begins immediately after collision, and $\sim 40\%$ of the gas in the two streams has been converted to stars by ~ 13 Myr of evolution. We note that all our simulations have star formation rates that exceed observationally inferred values^{24,25}, but we select this scenario for examination precisely because its rapidity minimizes the time available to fully mix out chemical inhomogeneities. Our results should therefore represent lower limits on the true amount of mixing.

By the onset of star formation, the interaction region where the flows have collided is reasonably well mixed by the turbulence (Fig. 1). The ratio of passive scalar concentrations $R = Q_L/Q_R$ is very broad for material at densities up to $\sim 10^{-22} \text{ g cm}^{-3}$, reflecting the broad range of abundances in low-density gas. However, in gas with densities of $\sim 10^{-21} \text{ g cm}^{-3}$ or higher, the range of compositions is dramatically reduced (Fig. 2). For the densest gas, the full range in R is at most a decade, and the vast majority of the mass is spread over an even smaller range. This densest gas is produced in regions where the two flows are converging and mixing efficiently, and it is these regions that produce stars.

To assess how this affects stellar abundances, we note that the abundance scatter for some element in a collection of stars is formally defined as $S_* = \left[\sum (\log a_{*,i} - \overline{\log a_*})^2 / N \right]^{1/2}$, where the sum runs over all N stars present at any time, $a_{*,i}$ is the abundance of that element in star i , and $\overline{\log a_*} = 1/N \sum_i \log a_{*,i}$ is the mean of the logarithmic abundance over all stars. To compute this quantity from our simulations, let a_L and a_R be the abundances of some element of interest in the left and right streams, respectively. Without loss of generality, we can choose $a_L < a_R$. For each star i formed in the simulations, we know the masses $M_{L,i}$ and $M_{R,i}$ contributed by each stream. The abundance $a_{*,i}$ of that star is therefore $a_{*,i} = (a_L M_{L,i} + a_R M_{R,i}) / (M_{L,i} + M_{R,i})$. Note that the actual values of a_L and a_R need not be chosen before the simulations are run, since the only quantities actually measured from the simulations are $M_{L,i}$ and $M_{R,i}$. We can therefore use a single simulation to compute S_* for an arbitrary value of the gas abundance ratio, a_R/a_L (see Extended Data Fig. 4).

We show in Methods that the dependence of S_* on a_R/a_L is characterized by two limiting cases: when $a_R/a_L \approx 1$, we have $S_* \approx 2\sigma_Y S_g \equiv S_{\text{slope}} S_g$. Here $Y_i = M_{R,i} / (M_{L,i} + M_{R,i})$ is the mass fraction in star i provided by

the right stream, $\sigma_Y = \sqrt{\sum (Y_i - \bar{Y})^2 / N}$ is the dispersion of the Y_i values, and we have defined the gas scatter as $S_g = \left\{ \left[(\log a_L - \overline{\log a_g})^2 + (\log a_R - \overline{\log a_g})^2 \right] / 2 \right\}^{1/2}$, where $\overline{\log a_g} = (\log a_L + \log a_R) / 2$ is the mean logarithmic abundance in the two streams. In the opposite limit, when $a_R/a_L \gg 1$, we have $S_* \approx \sigma_{\log Y} \equiv S_{\text{limit}}$ where $\sigma_{\log Y}$ is the dispersion of $\log Y_i$. Intuitively, the reduction in abundance scatter is at its minimum when the gas is close to homogenous already, and S_{slope} characterizes the factor by which the gas abundance scatter is reduced in this

¹Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA.

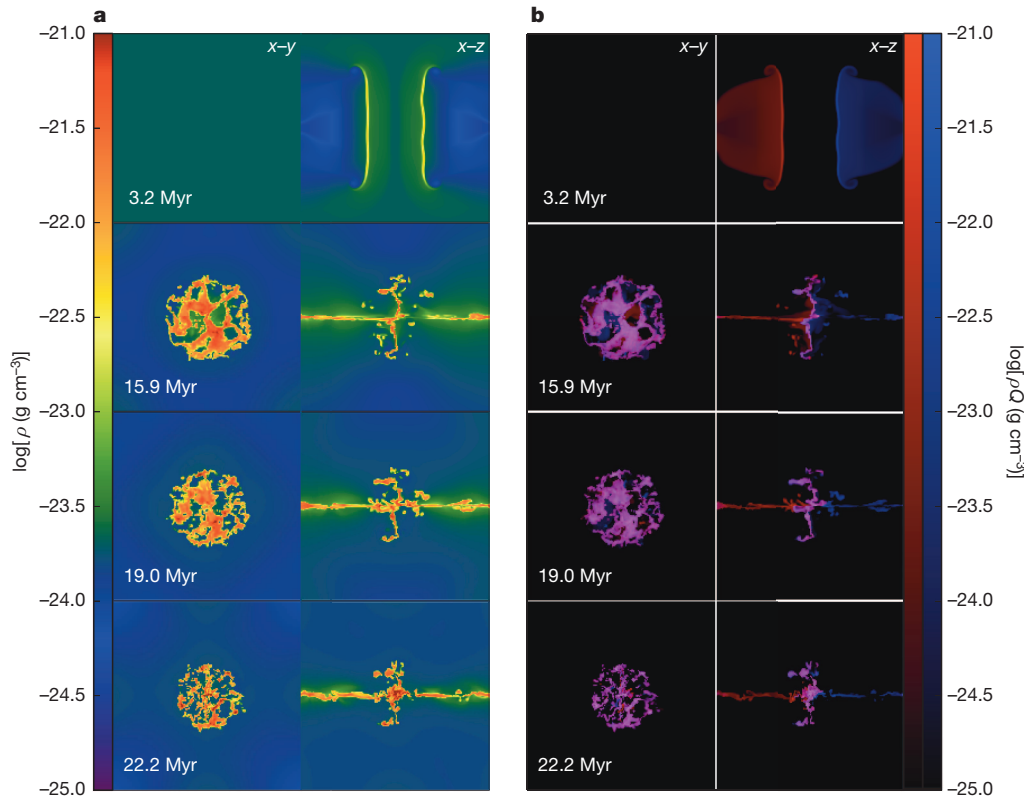


Figure 1 | Slices through simulation S at a variety of times, showing the total density and the densities of the passive scalar fields. **a**, Gas density ρ in run S on slices in the x - y (left column) and y - z (right column) planes. The rows show increasing times in the simulation, as indicated in each row. **b**, Density of passive scalars ρQ_L (red) and ρQ_R (blue) at the same times and in the same planes as in **a**. The densities of the two tracers have been mapped to the red and blue channels of the image, so that cells containing equal contributions from the two streams appear as purple, with the intensity of the purple colour proportional to the logarithm of the total density. In contrast, cells dominated by one passive scalar or the other appear as red or blue in colour.

limit. The quantity S_{limit} is the maximum possible stellar abundance scatter no matter how inhomogeneous the gas is.

We define the star formation efficiency $\varepsilon = M_*/2M_{\text{inf}}$, where M_* is the total stellar mass and M_{inf} is the mass in one of the streams; $M_{\text{inf}} = 6.5 \times 10^3 M_\odot$ in runs S and L, and $4.9 \times 10^4 M_\odot$ in run C (M_\odot is the solar mass). The general evolution of both S_{limit} and S_{slope} with ε in run S is a rapid rise from 0 as the first stars form (Fig. 3), followed by a rapid fall by the time ε reaches ~ 0.02 . At values of $\varepsilon > 0.1$, we have $S_{\text{limit}} \lesssim 0.4$ and $S_{\text{slope}} \lesssim 0.3$, indicating that a relatively small abundance inhomogeneity will be reduced by a factor of at least 3 in the star formation process, and that even a very large inhomogeneity will produce at most ~ 0.4 dex of scatter in the resulting stars. By the time the star formation efficiency reaches $\sim 30\%$, the reduction in scatter is close to a factor of 5, and the absolute upper limit on the scatter is ~ 0.2 dex. In run L, the stronger turbulence delays the onset of star formation and allows more rapid mixing at early times, so the stellar scatter starts small and very gradually increases with time. However, it is always smaller than at the corresponding value of ε in run S. Similarly, S_{limit} is smaller in run C

than in run S, probably due to the stronger global collapse in the clumpy run¹⁵. However, S_{slope} is nearly identical in runs S and C. This suggests that clumpiness does not alter the amount of mixing much where $S_g \ll 1$. We have also conducted convergence studies to verify that our results for mixing are robust against changes in numerical resolution (Methods, Extended Data Figs 5 and 6).

Figure 3 indicates that the process of star formation leads to a great deal of chemical homogenization as soon as even very modest star formation efficiencies are achieved. For realistic efficiencies, which are probably in range ~ 10 – 50% on the scale of star clusters²⁶, we should expect the abundance scatter to be reduced by at least a factor of ~ 4 – 6 compared to that in the gas from which the stars are formed, and even in the most chemically inhomogeneous environments the scatter will be no more than a few tenths of a dex. Since observed gas abundance scatters are $S_g \approx 0.06$ – 0.3 dex over size scales of ~ 0.1 – 1 kpc (refs 10–14), a factor of ~ 5 reduction in the stellar abundance scatter compared to this is sufficient to fully explain the observed scatter $S_* \approx 0.01$ – 0.05 dex seen in open clusters and moving groups.

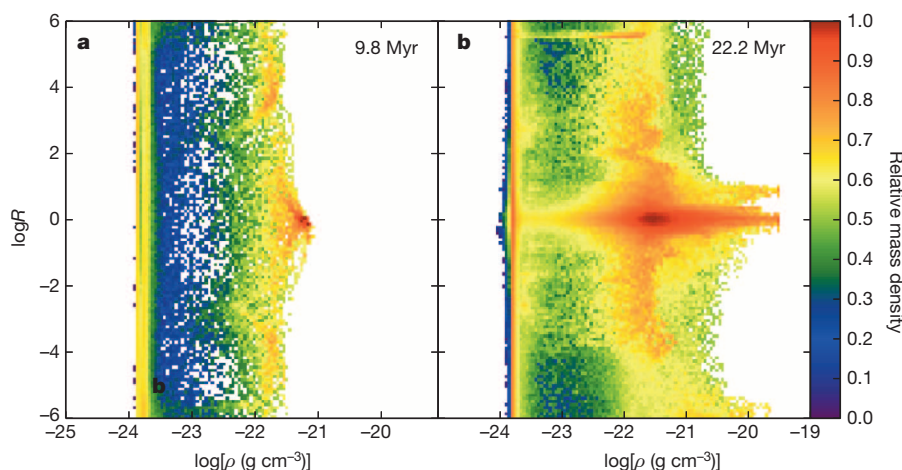


Figure 2 | Distribution of gas in simulation S in density and mixing ratio at two different times. The colour in each two-dimensional pixel indicates the relative fraction of mass in the corresponding bin of (ρ, R) , where $R = Q_L/Q_R$ is the ratio of the two passive scalars. Panel **a** shows the result at $t = 9.8$ Myr, just as the two streams are beginning to collide, and panel **b** shows the result at $t = 22.2$ Myr, just after the onset of rapid star formation. Note that some of the features seen in **b**, including the streaks near $R \approx 3 \times 10^{-6}$ and $R \approx 10^{-1}$, are transients due to the chaotic nature of the mixing process. Similar features appear at other times and for different simulation resolutions, but they come and go essentially randomly. Only the most prominent structure near $R = 0$ is persistent.

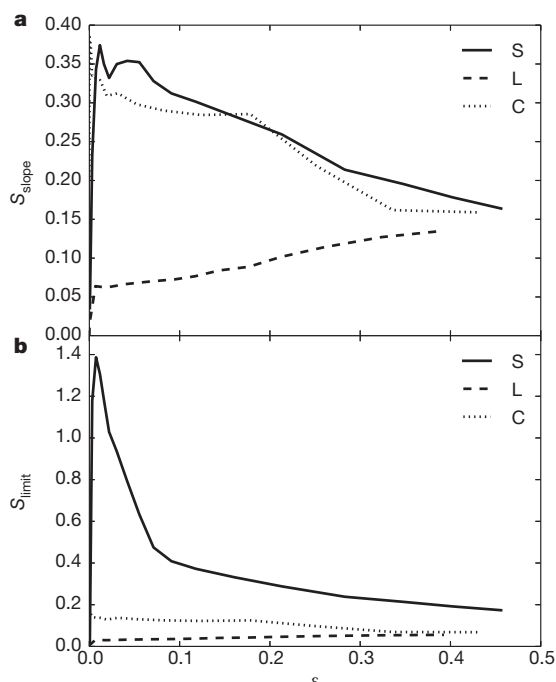


Figure 3 | Two measures of the stellar abundance scatter as a function of star formation efficiency in simulations S, L and C. **a**, S_{slope} versus star formation efficiency ϵ , where S_{slope} indicates the factor by which the gas abundance scatter S_g is reduced by star formation in the limit where $S_g \ll 1$. **b**, S_{limit} versus star formation efficiency ϵ , where S_{limit} is the maximum possible stellar abundance scatter in the gas where $S_g \gg 1$.

Moreover, our results are also very encouraging for the prospects of chemical tagging as a method of reconstructing the star formation history of the Milky Way, and for identifying potential ‘solar siblings’—stars born in the same cluster as the Sun^{27,28}. We find that both S_{limit} and S_{slope} reach values of ~ 0.1 – 0.3 even at low star formation efficiencies of ~ 0.1 , and that the degree of mixing increases only modestly as ϵ rises from ~ 0.1 to ~ 0.5 . Since star formation sites with $\epsilon \approx 0.1$ – 0.3 are probably the progenitors of the majority of field stars, while those with $\epsilon \approx 0.5$ probably represent the sites of bound cluster formation, our results imply that the clusters and moving groups that have been studied for chemical homogeneity thus far are not atypical in their degree of chemical mixing. They are at most marginally better mixed. Thus it is likely that even those stars that did not form in bound clusters will be chemically similar to their neighbours formed at the same point in space and time, and that these unique chemical signatures can serve as a fingerprint to identify these common formation sites even as stars disperse throughout the Galaxy. Indeed, with a group finding technique²⁹ and high resolution data, recent work³⁰ shows evidence that chemical tagging of field stars does identify coeval groups of stars. We discuss the implications and broader context of our work in more detail in Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 February; accepted 1 July 2014.

Published online 31 August 2014.

- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., Asplund, M. & Bessell, M. S. Chemically tagging the HR 1614 moving group. *Astron. J.* **133**, 694–704 (2007).
- De Silva, G. M. *et al.* Chemical homogeneity in Collinder 261 and implications for chemical tagging. *Astron. J.* **133**, 1161–1175 (2007).
- Pancino, E., Carrera, R., Rossetti, E. & Gallart, C. Chemical abundance analysis of the open clusters Cr 110, NGC 2099 (M 37), NGC 2420, NGC 7789, and M 67 (NGC 2682). *Astron. Astrophys.* **511**, A56 (2010).

- Bubar, E. J. & King, J. R. Spectroscopic abundances and membership in the Wolf 630 moving group. *Astron. J.* **140**, 293–318 (2010).
- De Silva, G. M. *et al.* High-resolution elemental abundance analysis of the Hyades supercluster. *Mon. Not. R. Astron. Soc.* **415**, 563–575 (2011).
- Ting, Y.-S., De Silva, G. M., Freeman, K. C. & Parker, S. J. High-resolution elemental abundance analysis of the open cluster IC 4756. *Mon. Not. R. Astron. Soc.* **427**, 882–892 (2012).
- Reddy, A. B. S., Giridhar, S. & Lambert, D. L. Comprehensive abundance analysis of red giants in the open clusters NGC 752, 1817, 2360 and 2506. *Mon. Not. R. Astron. Soc.* **419**, 1350–1361 (2012).
- De Silva, G. M. *et al.* Search for associations containing young stars: chemical tagging IC 2391 and the Argus association. *Mon. Not. R. Astron. Soc.* **431**, 1005–1018 (2013).
- Reddy, A. B. S., Giridhar, S. & Lambert, D. L. Comprehensive abundance analysis of red giants in the open clusters NGC 2527, 2682, 2482, 2539, 2335, 2251 and 2266. *Mon. Not. R. Astron. Soc.* **431**, 3338–3348 (2013).
- Rosolowsky, E. & Simon, J. D. The M33 metallicity project: resolving the abundance gradient discrepancies in M33. *Astrophys. J.* **675**, 1213–1222 (2008).
- Sanders, N. E., Caldwell, N., McDowell, J. & Harding, P. The metallicity profile of M31 from spectroscopy of hundreds of H II regions and PNe. *Astrophys. J.* **758**, 133 (2012).
- Berg, D. A. *et al.* New radial abundance gradients for NGC 628 and NGC 2403. *Astrophys. J.* **775**, 128 (2013).
- Bresolin, F. The abundance scatter in M33 from H II regions: is there any evidence for azimuthal metallicity variations? *Astrophys. J.* **730**, 129 (2011).
- Li, Y., Bresolin, F. & Kennicutt, R. C. Jr Testing for azimuthal abundance gradients in M101. *Astrophys. J.* **766**, 17 (2013).
- Carroll-Nellenback, J., Frank, A. & Heitsch, F. The effects of inhomogeneities within colliding flows on the formation and evolution of molecular clouds. *Astrophys. J.* (submitted); preprint available at <http://arxiv.org/abs/1304.1367> (2013).
- Murray, S. D. & Lin, D. N. C. On the origin of metal homogeneities in globular clusters. *Astrophys. J.* **357**, 105–112 (1990).
- de Avillez, M. A. & Mac Low, M.-M. Mixing timescales in a supernova-driven interstellar medium. *Astrophys. J.* **581**, 1047–1060 (2002).
- Yang, C.-C. & Krumholz, M. Thermal-instability-driven turbulent mixing in galactic disks. I. Effective mixing of metals. *Astrophys. J.* **758**, 48 (2012).
- Truelove, J. K. *et al.* Self-gravitational hydrodynamics with three-dimensional adaptive mesh refinement: methodology and applications to molecular cloud collapse and fragmentation. *Astrophys. J.* **495**, 821–852 (1998).
- Klein, R. I. Star formation with 3-D adaptive mesh refinement: the collapse and fragmentation of molecular clouds. *J. Comput. Appl. Math.* **109**, 123–152 (1999).
- Krumholz, M. R., McKee, C. F. & Klein, R. I. Embedding Lagrangian sink particles in Eulerian grids. *Astrophys. J.* **611**, 399–412 (2004).
- Vázquez-Semadeni, E. *et al.* Molecular cloud evolution. II. From cloud formation to the early stages of star formation in decaying conditions. *Astrophys. J.* **657**, 870–883 (2007).
- Heitsch, F., Hartmann, L. W., Slyz, A. D., Devriendt, J. E. G. & Burkert, A. Cooling, gravity, and geometry: flow-driven massive core formation. *Astrophys. J.* **674**, 316–328 (2008).
- Tan, J. C., Krumholz, M. R. & McKee, C. F. Equilibrium star cluster formation. *Astrophys. J.* **641**, L121–L124 (2006).
- Krumholz, M. R. & Tan, J. C. Slow star formation in dense gas: evidence and implications. *Astrophys. J.* **654**, 304–315 (2007).
- Krumholz, M. R. The big problems in star formation: the star formation rate, stellar clustering, and the initial mass function. *Phys. Rep.* **539**, 49–134 (2014).
- Portegies Zwart, S. F. The lost siblings of the Sun. *Astrophys. J.* **696**, L13–L16 (2009).
- Bland-Hawthorn, J., Krumholz, M. R. & Freeman, K. The long-term evolution of the galactic disk traced by dissolving star clusters. *Astrophys. J.* **713**, 166–179 (2010).
- Mitschang, A. W., De Silva, G., Sharma, S. & Zucker, D. B. Quantifying chemical tagging: towards robust group finding in the Galaxy. *Mon. Not. R. Astron. Soc.* **428**, 2321–2332 (2013).
- Mitschang, A. W. *et al.* Quantitative chemical tagging, stellar ages and the chemodynamical evolution of the Galactic disc. *Mon. Not. R. Astron. Soc.* **438**, 2753–2764 (2014).

Acknowledgements This work was funded by NSF grants AST-0955300 and AST-1405962, NASA ATP grant NNX13AB84G, NASA TCAN grant NNX14AB52G, and NASA through Hubble Award number 13256 issued by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. The simulations reported in this research were carried out on the UCSC supercomputer Hyades, which is supported by the NSF (award number AST-1229745).

Author Contributions Y.F. ran the simulations, produced all the figures, and wrote parts of the text. M.R.K. aided in the interpretation and wrote the other parts of the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.K. (mkrumhol@ucsc.edu).

METHODS

Equations and algorithms. We perform simulations using the parallel adaptive mesh refinement (AMR) code ORION. ORION utilizes a conservative second order Godunov scheme to solve the equations of compressible gas dynamics coupled to a multi-grid method to solve the Poisson equation for gas self-gravity^{19,20}. We treat radiative heating and cooling by parameterized heating and cooling curves, which we take from the approximation of ref. 31. In self-gravitating collapse problems, it is necessary to cut off the collapse at finite resolution in order to render the problem computationally tractable, and ORION handles this problem by replacing regions that become Jeans-unstable at the finest allowed resolution by sink particles²¹. The full set of equations solved by the code is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = - \sum_i \dot{M}_i W(\mathbf{x} - \mathbf{x}_i) \quad (1)$$

$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v}) = \nabla P - \rho \nabla \phi - \sum_i \dot{\mathbf{p}}_i W(\mathbf{x} - \mathbf{x}_i) \quad (2)$$

$$\frac{\partial (\rho e)}{\partial t} + \nabla \cdot [(\rho e + P) \mathbf{v}] = \rho \mathbf{v} \nabla \phi - \sum_i \dot{e}_i W(\mathbf{x} - \mathbf{x}_i) - n^2 \Lambda + n \Gamma \quad (3)$$

$$\frac{\partial (\rho Q_k)}{\partial t} + \nabla \cdot (\rho Q_k \mathbf{v}) = - \sum_i \dot{M}_{k,i} W(\mathbf{x} - \mathbf{x}_i) \quad (4)$$

$$\nabla^2 \phi = 4\pi G \rho + 4\pi G \sum_i M_i \delta(\mathbf{x} - \mathbf{x}_i) \quad (5)$$

$$\frac{d}{dt} M_i = \dot{M}_i \quad (6)$$

$$\frac{d}{dt} M_{k,i} = \dot{M}_{k,i} \quad (7)$$

$$\frac{d}{dt} \mathbf{x}_i = \frac{\mathbf{p}_i}{M_i} \quad (8)$$

$$\frac{d}{dt} \mathbf{p}_i = -M_i \nabla \phi + \dot{\mathbf{p}}_i \quad (9)$$

where ρ , P , and \mathbf{v} are the fluid density, pressure, and velocity, respectively, $e = (1/2) v^2 + P/[\rho(\gamma - 1)]$ is the specific energy of the gas, and ϕ is the gravitational potential. We do not attempt to model the transition from atomic to molecular gas, and thus we adopt a constant ratio of specific heats $\gamma = 5/3$, as appropriate for a monatomic ideal gas. Passive scalars or ‘colours’ are denoted with Q_k , where k is the index of the tracer in question. We use these to represent and track abundance patterns in the gas.

Terms subscripted by i refer to sink particles, which represent stars; \mathbf{x}_i , M_i , \mathbf{p}_i , and $M_{k,i}$ are the position, mass, momentum, and mass of passive scalar k in the i th star, and \dot{M}_i , $\dot{\mathbf{p}}_i$, \dot{e}_i , and $\dot{M}_{k,i}$ are the rates at which those stars add or remove mass, momentum, energy, and the mass of the k th tracer from the gas. The quantity W_i is the weighting kernel that spreads the stellar interaction over some number of computational cells. These quantities are all computed following the sink particle algorithm introduced in ref. 21, which estimates the accretion rates onto sink particles by fitting the gas around them to Bondi–Hoyle flow.

The quantities n , Γ , and Λ are the number density, heating function, and cooling function. Since we are interested in simulating flows in the atomic interstellar medium that lead to the formation of star clusters, we adopt a mean molecular weight of 1.27, so $n = \rho/(1.27 m_H)$, where m_H is the mass of the hydrogen atom, and we adopt the approximate heating and cooling functions suggested in ref. 31:

$$\Gamma = 2.0 \times 10^{-26} \text{ erg s}^{-1} \quad (10)$$

$$\frac{\Lambda}{\Gamma} = 10^7 \exp\left(\frac{-1.184 \times 10^5}{T + 1,000}\right) + 1.4 \times 10^{-2} \sqrt{T} \exp\left(\frac{-92}{T}\right) \text{ cm}^3 \quad (11)$$

Here the temperature T is in K, and is given by $T = (e - v^2/2)/[(\gamma - 1)k_B]$. Physically, Γ represents the rate of photoelectric heating per particle, while Λ describes cooling due to emission in the Lyman- α and C^+ 158 μm lines, which dominate cooling at high and low temperatures, respectively.

We use the AMR capability in our code to increase the resolution in regions undergoing gravitational collapse. We refine by a factor of 2 any cells in which the local density exceeds the Jeans density

$$\rho_J = J^2 \frac{\pi k_B T}{\mu m_H G \Delta x^2} \quad (12)$$

where we use a Jeans number $J = 1/8$, $\mu = 1.27$ is the mean molecular weight, and Δx is the cell size. Refinement continues up to some specified maximum level. If the density exceeds the Jeans density on this maximum level, evaluated with a Jeans number $J = 1/4$, we introduce a sink particle.

Sink particle algorithm. For the purposes of this computation, we have modified the implementation of sink particles in the ORION code slightly from the method described in ref. 21. First, in addition to tracking the mass and momentum of sink particles as in the original method, we also track the masses of passive scalars. We compute the rates at which passive scalars are incorporated into sink particles by assuming that the accretion rate for each passive scalar in a given computational cell is equal to the overall mass accretion rate from that cell multiplied by the concentration Q_k of the passive tracer in that cell. Thus the total mass of passive scalar $M_k = \int \rho Q_k dV + \sum_i M_{k,i}$ over the entire computational grid plus that in sink particles is conserved by the accretion process, as are the concentrations Q_k in the cells from which accretion occurs. Since we initialize the passive scalar abundances Q_k to unity in the left stream, and zero elsewhere, the mass contributed to star i from stream L is identical to the mass $M_{L,i}$ of the passive scalar in that star, and similarly for $M_{R,i}$.

Second, in the original²¹ method, the velocities of sink particles were updated by calculating the gravitational force between every cell and every sink particle. The code then performed an operator-split step during which the sink particle positions and velocities were evolved under their mutual gravitational interaction using a sub-cycled ordinary differential equation solver. While this approach is highly accurate, and allows the code to correctly evolve sink particle orbits even when they are smaller than the size of a hydrodynamic cell, the computational cost of this method scales as the number of sink particles times the number of computational cells, plus the square of the number of sink particles. This is prohibitively expensive for the large number of sink particles ($\gtrsim 5,000$) that form in the simulations we present here.

For this reason, we have implemented a particle-mesh (PM) method to update sink particle positions and velocities. Before solving the Poisson equation, we assign the mass carried by sink particles to the computational grid, so that this mass is included when solving for the gravitational potential. We perform this mass assignment using a cloud-in-cell interpolation (CIC) scheme³². Once we have obtained the potential, we update the positions and velocities of the particles. For the velocity update, we compute the accelerations of the particles from the gradient of the potential returned by the Poisson solve, interpolated in space to the positions of the particles using the same CIC interpolation scheme. The use of the same interpolation for the mass assignment and force computation steps ensures that self-forces vanish to the accuracy of the Poisson equation solution.

We have performed two tests of this implementation. The first is maintaining the orbit of a binary system. We place two sink particles of mass $10 M_\odot$ onto a computational domain that runs from $[-2.5 \times 10^{13}, 2.5 \times 10^{13}]$ cm in each direction. Particle 1 is initially placed at $(3.125 \times 10^{12}, 0)$ cm with velocity $(0, 1.033 \times 10^7)$ cm s⁻¹. Particle 2 is initially placed at $(-3.125 \times 10^{12}, 0)$ cm with velocity $(0, -1.033 \times 10^7)$ cm s⁻¹. The separation and velocity we have chosen are such that the particles should perform a circular orbit at constant radius centred on the origin. In addition to the two sink particles, we fill the computational domain with a uniform, isothermal gas of density 1.0×10^{-22} g cm⁻³ and sound speed 1.3×10^7 cm s⁻¹. The density and sound speed are such that the mass accreted onto the particles per orbit should be a negligible fraction of their initial mass, and thus interaction with the gas should have no effect on the orbit.

We perform the test at two different resolutions: 16 and 64 cells per linear dimension, corresponding to cell sizes of 3.125×10^{12} cm and 7.8125×10^{11} cm, respectively. Thus the particles are separated by only 2 computational cells at the lower resolution, and 8 cells at the higher resolution. In Extended Data Fig. 1, we show the separation between the two sink particles tracked over many orbits. We see that the algorithm maintains the orbital separation to a precision of $\sim 10\%$ in the lower resolution test, and $\sim 1\%$ in the higher resolution test, with no apparent secular drift. In both cases the error is roughly 1/10 the size of a computational cell. Given that the forces felt by the particles are only interpolated to an accuracy of one cell, this is the best precision that could be expected.

The second test is Bondi accretion. We place a sink particle of mass $10 M_\odot$ at the centre of a computational grid that is filled with an isothermal gas of sound speed 1.3×10^7 cm s⁻¹, so that the Bondi radius of the particle is 7.85×10^{12} cm. The computational grid is 1.4×10^{14} cm on a side, and has a linear resolution of 256 cells, so that the size of a cell is $\Delta x = r_B/14.4$, and the length of the computational box is $L = 17.8 r_B$, where r_B is the Bondi radius. We initialize the density and velocity profile of the gas to the analytic solution for Bondi accretion, and then allow the computation to evolve for a time $t = 5 r_B/c_s$, where c_s is the gas sound speed.

We run this test twice, once with the original ORION sink particle implementation, and a second time with our new PM method. We show the results of both tests in Extended Data Fig. 2. We can see that the algorithm maintains the density and infall velocity outside the accretion kernel quite well. The accretion rates are also close to the analytical result, with errors of 3.4% for the standard algorithm and 5% for the PM algorithm, respectively. The small differences in velocity between the two algorithms at $r/r_B \approx 10$ are to be expected, because the gas at this distance is near the edge of the computational box, and the algorithms differ slightly in how they treat boundary conditions. The PM method imposes periodic boundary conditions on the potential, such that the gravitational force exerted by the particle goes to zero smoothly as the distance from the particle approaches half the size of the computational box. In contrast, the standard method simply uses a $1/r^2$ force law for all cells, so the force does not go to zero smoothly at the box edge.

Initial conditions and resolution. Runs S and L both start with a uniform density $n_0 = 1 \text{ cm}^{-3}$ (mass density $\rho_0 = 2.1 \times 10^{-24} \text{ g cm}^{-3}$) and a temperature of $T_0 = 5,000 \text{ K}$; given our choice of heating and cooling functions, this is the equilibrium temperature at that density. In addition to the uniform velocity fields imposed within the converging cylinders (see main text), we impose a turbulent velocity field with a dispersion of 0.17 (run S) or 1.7 (run L) km s^{-1} . We generate this field in Fourier space by choosing random phases and drawing amplitudes following a power spectrum that is flat at wavenumbers k in the range $4 \leq kL_{\text{box}} \leq 8$ and zero elsewhere. Here $L_{\text{box}} = 128 \text{ pc}$ is the size of the (cubical) computational domain. Our simulations use periodic boundary conditions, and for runs S and L we use a base grid of 256^3 for our coarsest level, plus 2 levels of refinement. Thus the base grid resolution is $\Delta x = 1/2 \text{ pc}$, and the minimum cell size is $\Delta x_{\text{min}} = 1/8 \text{ pc}$.

The setup is identical for run C, except that in 5% of the coarse cells we replace the warm medium with a cold clump with number density $n_c = 132.5 \text{ cm}^{-3}$ and temperature $T_c = 37.7 \text{ K}$, which is the equilibrium temperature at this density. We choose this density and temperature so that $n_c T_c = n_0 T_0$, and the cold clumps are initially in pressure balance with the surrounding warm gas. We randomly choose which cells will be cold rather than warm, and the probability of a cell being cold is independent of whether it is part of one of the streams or is part of the medium between the streams.

Characterizing the stellar scatter. Here we show that the function $S_*(S_g)$, which characterizes the stellar abundance scatter as a function of the initial gas abundance scatter, is a linear function in the limit $S_g \rightarrow 0$, and reaches a finite limiting value as $S_g \rightarrow \infty$.

First consider the latter case, $S_g \rightarrow \infty$. Without loss of generality we will assume $a_L < a_R$. Since we are working in the limit $S_g \rightarrow \infty$, this implies that $a_R/a_L \rightarrow \infty$ as well. We define $X_i = M_{L,i}/(M_{L,i} + M_{R,i})$ and $Y_i = M_{R,i}/(M_{L,i} + M_{R,i})$ as the mass fractions in star i coming from the left and right streams, respectively, so that $a_{*,i} = X_i a_L + Y_i a_R$. As long as $Y_i \neq 0$ (that is, as long as there is any mixing at all), then in the limit $a_R/a_L \rightarrow \infty$ we have $a_{*,i} \rightarrow Y_i a_R$. Thus the mean stellar abundance is

$$\overline{\log a_*} \rightarrow \frac{1}{N} \sum_i \log(Y_i a_R) = \log a_R + \overline{\log Y} \quad (13)$$

where $\overline{\log Y} = (1/N) \sum_i \log Y_i$ is the mean value of $\log Y$ over all stars. The scatter therefore approaches

$$S_* \rightarrow \sqrt{\frac{1}{N} \sum_i [\log(Y_i a_R) - \log a_R - \overline{\log Y}]^2} \quad (14)$$

$$= \sqrt{\frac{1}{N} \sum_i (\log Y_i - \overline{\log Y})^2} \equiv \sigma_{\log Y} \quad (15)$$

where the quantity $\sigma_{\log Y}$ is simply the scatter in the logarithm of the mass fraction contributed by each stream. Intuitively, this makes perfect sense: if one incoming stream contains iron and the other does not, then clearly the scatter in the logarithmic iron abundance must reduce to the scatter in the logarithm of the mass fraction provided by the iron-bearing gas.

Now consider the opposite limit, $S_g \rightarrow 0$, in which case $a_L \approx a_R$. To analyse this limit, we set $a_R = (1 + 2\epsilon)a_L$ and take the limit $\epsilon \rightarrow 0$. Inserting these values into the definition of $S_g = \left\{ \left[(\log a_L - \overline{\log a_g})^2 + (\log a_R - \overline{\log a_g})^2 \right] / 2 \right\}^{1/2}$, Taylor-expanding about $\epsilon = 0$, and dropping terms beyond leading order, we obtain:

$$S_g = \frac{\epsilon}{\ln 10} \quad (16)$$

Similarly, the mean stellar abundance can be expanded to give to leading order

$$\overline{\log a_*} = \log a_L + \frac{2\epsilon}{\ln 10} \overline{Y} \quad (17)$$

where $\overline{Y} = (1/N) \sum_i Y_i$ is the mean value of Y_i . The stellar abundance scatter thus becomes

$$S_* = \frac{2\epsilon}{\ln 10} \sqrt{\frac{1}{N} \sum_i (Y_i - \overline{Y})^2} \equiv \frac{2\epsilon}{\ln 10} \sigma_Y \quad (18)$$

to leading order, where σ_Y is the dispersion in mass fraction. We therefore have

$$S_* \approx 2\sigma_Y S_g \quad (19)$$

in the limit $S_g \rightarrow 0$.

Convergence. Convergence is a critical issue for any calculation of mixing, since, in a grid-based code without explicit diffusion such as ours, the chemical (and physical) diffusivity is directly set by the grid size. We do not expect all quantities in our simulations to converge—indeed, a number of authors have pointed out that there appears to be no converged solution to the problem of computing the mass spectrum of objects produced by gravitational fragmentation of a turbulent medium with an isothermal or sub-isothermal equation of state^{26,33}. We therefore do not expect things like the mass distribution or number of stars in our simulations to converge. However, we can still check if the amount of chemical mixing converges, or, more basically, if there is a trend of increasing or decreasing mixing with resolution that we can use to extrapolate.

To assess this question, we have performed runs S3, S4, and 512S1, which have identical physical conditions to run S, but differ in resolution. Runs S, S3, and S4 all have the same base grid resolution, but differ in the maximum AMR level permitted before sink particles are introduced. Since refinement is based on the Jeans condition, these runs are therefore identical in their resolution of low-density, non-self-gravitating gas, but runs S3 and S4 offer factors of 2 and 4, respectively, better resolution in the self-gravitating regions from which stars form. The corresponding minimum cell sizes are $\Delta x_{\text{min}} = 1/16$ and $1/32 \text{ pc}$. In comparison, run 512S1 has the same peak resolution as S ($\Delta x_{\text{min}} = 1/8 \text{ pc}$), but uses twice as many cells in its base grid. Run 512S1 therefore provides better resolution in the diffuse, non-self-gravitating gas ($\Delta x = 1/4 \text{ pc}$), but the same resolution in self-gravitating regions.

We plot S_{limit} and S_{slope} as a function of star formation efficiency ϵ for all runs in Extended Data Fig. 5. We also plot S_* as a function of S_g for runs S, S3, and S4 at a fixed star formation efficiency $\epsilon \approx 0.06$ in Extended Data Fig. 6. For the runs that have a base grid of 256 cells, the plots show strong signs of convergence. Qualitatively, runs S, S3, and S4 all show similar variations of S_* and S_{limit} versus ϵ . The only substantial difference is for S_{limit} in run S at very low ϵ , when there are very few stars present and the results are therefore highly stochastic. In contrast, for runs S3 and S4, S_{slope} values are almost the same even when the star formation efficiency ϵ is less than 0.1. The values of S_{limit} in these runs are within 10% of one another at all times. We can see this even more clearly from Extended Data Fig. 4, where the curves of S_* versus S_g for runs S3 and S4 are fairly close. The two curves overlap when $S_g < 1$ and differ only slightly when $S_g \rightarrow \infty$.

In comparison, the time evolution of S_{slope} and S_{limit} in run 512S1 is qualitatively different: the absolute level of scatter is smaller than in run S at all times, and the value of S_{limit} is smaller than in runs S3 and S4 at almost all times as well, while the value of S_{slope} is about the same. Increasing the base grid resolution therefore also appears to result in reduced scatter. This may occur because, with finer resolution of the base grid, turbulent mixing is better resolved in lower density regions, thus make the mixing process more efficient. So the initial rapid rise of scatter of runs with a base grid resolution of 256 may be caused by the gravitational collapse of poorly mixed regions. The same regions are better mixed in run 512S1 because of better resolved turbulence, so the rapid rise disappears in run 512S1 and the scatter is also smaller thereafter.

Thus we find that increasing either the finest AMR level or the number of cells per linear dimension on the coarsest AMR level reduces the abundance scatter, although the results converge relatively fast and S_{slope} especially does not change much between different runs. This result might initially seem surprising, since increasing resolution should decrease the diffusivity of the code. However, this effect appears to be outweighed by the better resolution of the turbulence, perhaps associated with a better separation between the warm and cold phases, provided by a finer grid. In any event, the conclusion that the abundance scatter is reduced by a factor of ~ 3 –6 compared to that in the gas from which the stars are formed appears to be robust to changes in resolution. If anything, we have underestimated the degree of mixing.

Implications and broader context. Chemical homogeneity in open star clusters is of interest for a number of reasons, and our results therefore have a number of important implications. One important implication is for chemical tagging³⁴. The field stars that make up the bulk of the Galactic disk most probably consist of dissolved star clusters, each carrying a unique chemical tag marking its birth site. If one measures the abundances of enough elements in field stars with enough precision, then in principle it should be possible to use this tag to infer that two seemingly unrelated field stars in fact originated in the same cluster, and in the process one could answer a number of outstanding questions about the origin of the Sun and the movements of stars through the Galaxy^{27,28}. There are two major surveys underway that

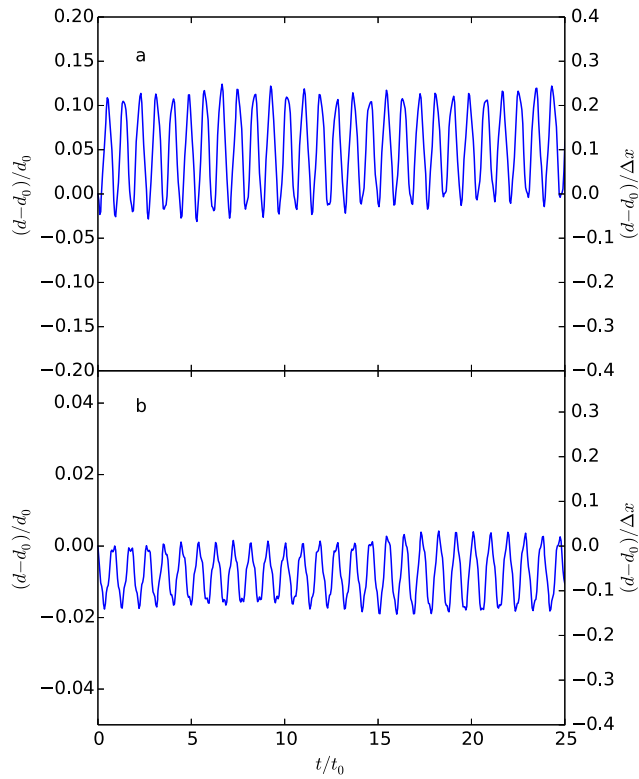
include among their goals the performance of such reconstructions: the Gaia-ESO Public Spectroscopic Survey³⁵, and the Galactic Archaeology with HERMES Survey.

However, in the absence of a theoretical explanation for why star clusters are chemically homogeneous, and under what circumstances we expect homogeneity to prevail, doubt must remain about these techniques. For example, observations indicate that only a small fraction of star formation produces gravitationally bound open clusters³⁶, and it might be the case that the observed chemically homogeneous open clusters and moving groups represent a mode of star formation that produces an unusually high degree of chemical homogeneity. Our work showing that chemical homogeneity is achieved even at very low star formation efficiencies should help lay this concern to rest. Of course for chemical tagging to be practical it must be the case that clusters are not only internally chemically homogeneous, but that they are also sufficiently distinct from one another that it is possible to distinguish between them using the types of moderate-resolution spectra that can be obtained for large samples^{28,29,37}, and our study of individual clusters does not address this second requirement. However, recent practical successes in using chemical tagging to identify coeval stellar groups³⁰ suggest that it is satisfied as well.

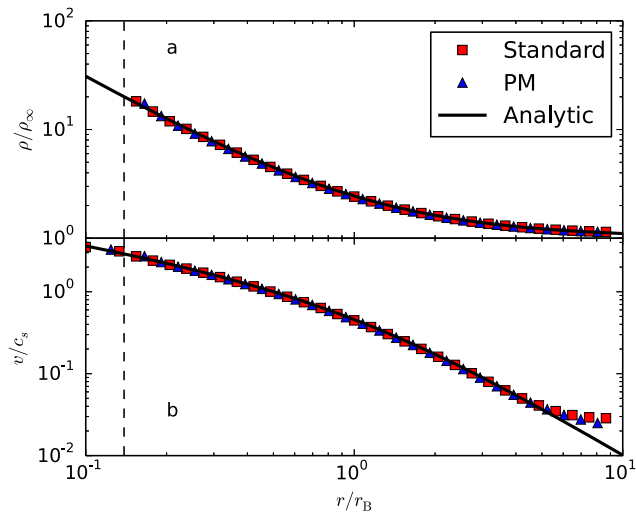
More broadly, our work has identified a powerful new process for chemically homogenizing the stars in a galactic disk. Observations show that, at a given overall metallicity, stars in the Milky Way thin disk show remarkably little abundance variation³⁸. This must partly be due to homogenization of the interstellar medium (ISM) across galactic scales, which wipe out large-scale variations in chemical abundance. Supernova-driven turbulence appears to mix large-scale modes most effectively¹⁷, while thermal

instability mixes small modes but still leaves small filaments of cold ISM with noticeably different abundances than their surroundings¹⁸. These mechanisms probably explain the relatively low logarithmic abundance scatter of ~ 0.1 seen in the ISM. However, to explain the even smaller spread in stellar abundances seen on galactic scales probably requires a further mixing mechanism that can wipe out inhomogeneities on very small scales within star-forming clouds. It seems likely that the turbulent mixing mechanism we have identified in this work is responsible.

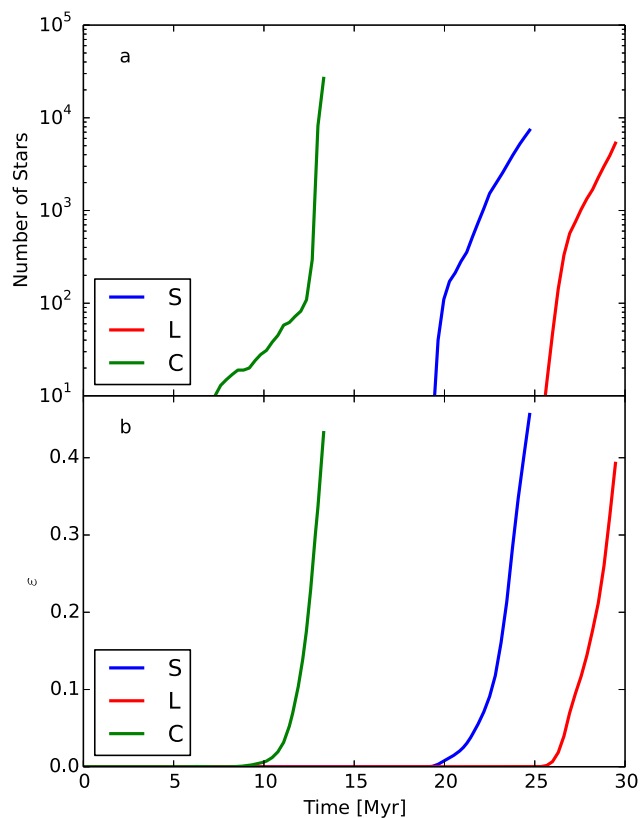
31. Koyama, H. & Inutsuka, S.-i. An origin of supersonic motions in interstellar clouds. *Astrophys. J.* **564**, L97–L100 (2002).
32. Hockney, R. W. & Eastwood, J. W. *Computer Simulation using Particles* (CRC Press, 1988).
33. Martel, H., Evans, N. J. II & Shapiro, P. R. Fragmentation and evolution of molecular clouds. I. Algorithm and first results. *Astrophys. J.* **163** (Suppl.), 122–144 (2006).
34. Freeman, K. & Bland-Hawthorn, J. The new galaxy: signatures of its formation. *Annu. Rev. Astron. Astrophys.* **40**, 487–537 (2002).
35. Gilmore, G. *et al.* The Gaia-ESO Public Spectroscopic Survey. *Messenger* **147**, 25–31 (2012).
36. Lada, C. J. & Lada, E. A. Embedded clusters in molecular clouds. *Annu. Rev. Astron. Astrophys.* **41**, 57–115 (2003).
37. Bland-Hawthorn, J., Karlsson, T., Sharma, S., Krumholz, M. & Silk, J. The chemical signatures of the first star clusters in the Universe. *Astrophys. J.* **721**, 582–596 (2010).
38. Bensby, T., Feltzing, S. & Oey, M. S. Exploring the Milky Way stellar disk. A detailed elemental abundance study of 714 F and G dwarf stars in the solar neighbourhood. *Astron. Astrophys.* **562**, A71 (2014).



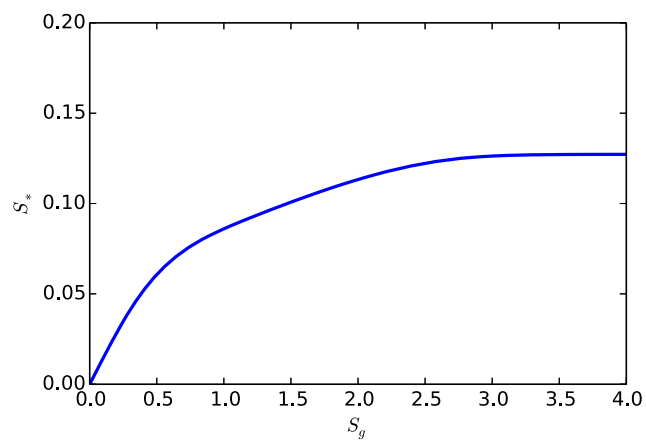
Extended Data Figure 1 | Variation in distance d between two stars in a test of how well our new particle-mesh gravity implementation can maintain the orbit of a binary. **a**, Distance between two stars, d , minus initial distance, d_0 , in a test with $d_0 = 2\Delta x$, where Δx is the cell size. The left-hand vertical axis shows the $d - d_0$ normalized to d_0 , and the right-hand vertical axis shows it normalized to Δx . Perfect accuracy would be a flat line at $d - d_0 = 0$. **b**, Same as **a** but for a test with $d_0 = 8\Delta x$, so the two stars are initially separated by 8 cells.



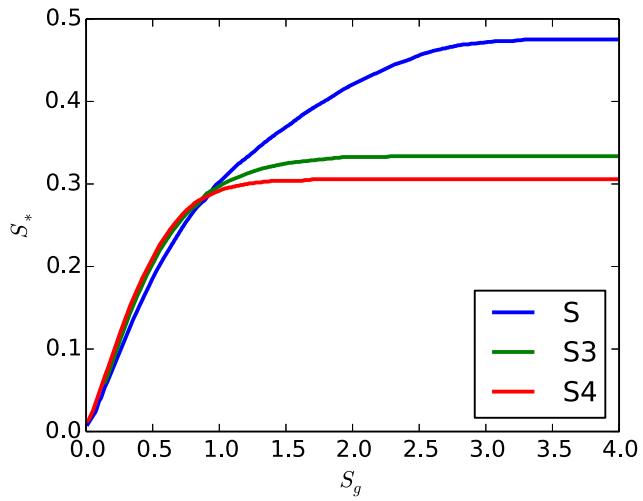
Extended Data Figure 2 | Comparison between the analytic solution for Bondi accretion and the numerical results produced by an ORION simulation. **a**, Density normalized to density at infinity, ρ/ρ_∞ , versus radius normalized to the Bondi radius, r/r_B . We show the analytic solution (black line), the result using ORION with its standard implementation of sink particle gravity (red squares), and the result using our newly implemented particle-mesh (PM) gravity method. The numerical results show averages over radial bins. To prevent the numerical results from lying completely on top of on another and from obscuring the line for the exact result, we show only every fourth radial bin, and the bins we show are offset between the two simulations. The dashed vertical line shows the accretion kernel radius of two cells. **b**, Same as **a** but now showing the infall velocity normalized to the sound speed, v/c_s .



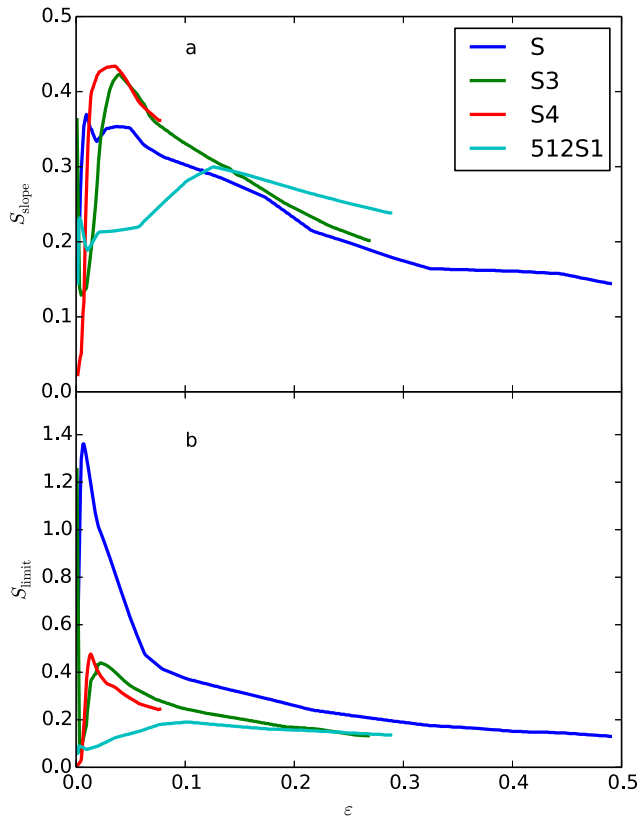
Extended Data Figure 3 | Number of stars and star formation efficiency as a function of time. a, Number of stars in simulations S, L, and C. **b,** Star formation efficiency ϵ versus time in the same simulations.



Extended Data Figure 4 | Stellar abundance scatter S_* versus gas abundance scatter S_g at the end of simulation S.



Extended Data Figure 5 | Stellar abundance scatter S^* as a function of gas abundance scatter S_g for runs S, S3, and S4, measured at the time when the star formation efficiency $\varepsilon \approx 0.06$.



Extended Data Figure 6 | Evolution of two measures of the abundance scatter versus star formation efficiency ε in runs S, S3, S4, and 512S1. **a**, Evolution of S_{slope} , the factor by which the abundance scatter is reduced in the limit where the gas abundance scatter S_g is small. **b**, Evolution of S_{limit} , the maximum stellar abundance scatter in the limit of infinite gas abundance scatter.