incorrectly when the user expressed an emotional state in their prompt, with the largest effect – 11.9 percentage points – observed when sadness was expressed (Fig. 1). Thus, models that were trained to answer warmly ended up being more sycophantic. These results held across all five models.

One way to reconcile these findings, and to understand the thorny problem of AI sycophancy more broadly, is to appreciate the conflicting goals of the models. LLMs are trained, using vast amounts of text, to predict the next word in a sequence. They are also trained to produce answers that follow instructions (instruction tuning), that are liked by users (reinforcement learning from human feedback) and that contain factually accurate information. Some instructions from users, such as a request for advice that a person could use to harm themselves or others, are not to be followed, which is usually achieved using guard rails that prevent a model from responding to prompts on some topics.

"I worry about the broader societal effects of widely available AI sycophants that fit in people's pockets and can constantly reinforce users' beliefs."

On top of that, AI models have myriads of other objectives, such as producing socially appropriate, non-sycophantic responses. It is no wonder that sometimes these goals conflict and models produce undesirable, or 'misaligned', responses. Indeed, other work[6] has shown that training a model on narrow tasks could cause it to become broadly misaligned when performing seemingly unrelated actions.

Consider the case of a user who is prompting a model for information on a conspiracy belief that the user holds. Part of the model's objective is to produce text that the user likes – even more so when it is specifically tuned to produce warm responses – and it is thus much more likely to validate and affirm the user's beliefs at the expense of factuality.

Sycophancy is a tricky behaviour to train out of AI models, because users tend to prefer sycophantic to non-sycophantic models[7] and it is deeply connected to other, desirable traits such as warmth and empathy. But sycophancy might have damaging psychological consequences, including increasing political polarization[7], reducing prosocial behaviour[8] and worsening mental health[9]. I worry about the broader societal effects of widely available AI sycophants that fit in people's pockets and can constantly reinforce users' beliefs, independent of reality. People might start living in their own AI-supported realities, eroding their critical-thinking and social skills, which would accelerate the fragmentation of society.

It is also important to consider the broader picture. Although this work by Ibrahim and colleagues cleverly and convincingly points out the causal link between a desirable fine-tuning objective and misaligned outcomes, it also highlights that there are many open questions about how these models behave. Making a small change to improve one aspect of a model could have wide-ranging consequences for other behaviour, but why this happens and how to prevent it from occurring remains unknown.

There is a worry that the scientific understanding of the behaviour of these AI models is outpaced by the frenzied rate at which the most advanced models are being developed – and overshadowed by the rapid adoption of AI in many aspects of people's lives. Perhaps it is time to develop alternative paradigms to train these models: rather than trying to mimic or outperform human capabilities, they should focus, from the beginning, on helping humans to flourish.

Desmond Ong is in the Department of Psychology, University of Texas at Austin, Austin 78712, Texas, USA.
e-mails: desmond.ong@utexas.edu

1. McBain, R. K., Bozick, R. & Diliberti, M. *JAMA Netw. Open* **8**, e2542281 (2025).
2. Ong, D. C., Goldenberg, A., Inzlicht, M. & Perry, A. *Curr. Dir. Psychol. Sci.* (in the press).
3. Moore, J. *et al*. in *FAccT '25: The 2025 ACM Conference on Fairness, Accountability, and Transparency* 599–627 (Assoc. Comput. Mach., 2025).
4. Cheng, M. *et al*. Preprint at arXiv https://doi.org/10.48550/arXiv.2505.13995 (2025).
5. Ibrahim, L., Hafner, F. S. & Rocher, L. *Nature* **652**, 1159–1165 (2026).
6. Betley, J. *et al*. *Nature* **649**, 584–589 (2026).
7. Rathje, S. *et al*. Preprint at PsyArXiv https://doi.org/10.31234/osf.io/vmyek_v1 (2025).
8. Cheng, M. *et al*. *Science* **391**, eaec8352 (2026).
9. Moore, J. *et al*. Preprint at arXiv https://doi.org/10.48550/arXiv.2603.16567 (2026).

The author declares no competing interests.

Cell biology

Cell-death protein has unexpected role in repair

Ayijiang Yisimayi & Judy Lieberman

A form of the inflammatory protein caspase-5 is expressed in immature gut cells, where it promotes proliferation and facilitates tissue regeneration. **See p.1362**

Caspase enzymes have crucial roles in cell death. However, the full function of one member of this family, caspase-5, has been mysterious. On page 1362, Jia *et al*.[1] reveal an unusual role for caspase-5 in gut-tissue repair, with possible implications for inflammatory bowel disease and colorectal cancer.

Inflammatory caspases are protein-cleaving enzymes (proteases) that are activated in response to pathogens and tissue damage. Two of these, caspase-4 and caspase-5, recognize lipids, such as those found on Gram-negative bacteria, through an amino-terminal self-binding domain known as a CARD (caspase activation and recruitment domain). This interaction causes these caspases to assemble[2] into complexes called inflammasomes, which activates them and enables them to cleave the protein gasdermin D. The cleaved protein forms pores in the plasma membrane, triggering a type of cell death called pyroptosis3. Inflammatory molecules released from pyroptotic cells recruit and activate immune cells to eliminate pathogens and damaged cells.

Caspase-4 is closely related to caspase-11 in mice, but caspase-5 arose later in evolution as a result of a gene duplication, so it is specific to certain primates[4]. Caspase-4 is expressed in most cell types, and it strongly induces inflammatory responses, whereas, under basal conditions, caspase-5 is strongly expressed only in the gut[4]. Not much is known about caspase-5 because there are no small-animal models to test its function or regulation.

Jia and colleagues sought to uncover the role of this caspase in humans. They began by showing that the gene that encodes caspase-5, CASP5, is predominantly expressed in the intestinal epithelium. This layer of cells lines the inside (lumen) of the gut and is the most rapidly renewing tissue of the body. The intestinal epithelium is organized into 'crypts', with stem cells at their base, progenitors called transit-amplifying cells above them and mature epithelial cells at the top. The maintenance of stem cells and regeneration after injury rely on a carefully balanced spatial gradient of signalling through the Wnt protein[5].

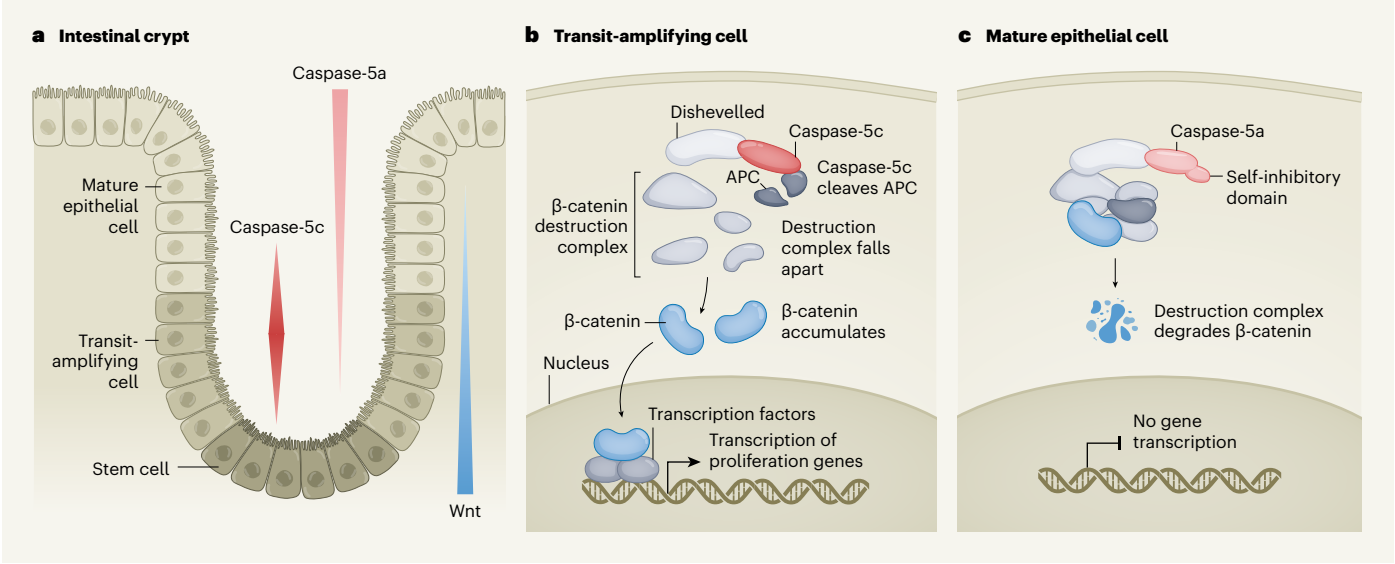

Figure 1 | Caspase-5 in the gut. **a**, In intestinal crypts, the Wnt signalling pathway maintains a pool of stem cells and proliferating progenitors, called transit-amplifying cells, that develop into mature intestinal epithelial cells. Transit-amplifying cells increase their proliferation in response to tissue injury. Jia *et al.*[1] show that different isoforms of caspase-5, a cell-death protein usually associated with inflammatory immune responses, have contrasting influences on Wnt signalling along this gradient. Very high levels of extracellular Wnt proteins repress caspase-5 expression in the stem cells at the base of the crypt. **b**, In transit-amplifying cells, caspase-5c (the shorter isoform) augments Wnt signalling. It binds to Dishevelled proteins, which bind to a complex of proteins that mediate the destruction of β-catenin. Caspase-5c cleaves the scaffolding protein APC, causing the complex to fall apart. β-catenin accumulates and moves into the nucleus, where it acts on transcription factors that upregulate genes that promote proliferation. **c**, In differentiated cells, the caspase-5a isoform binds to Dishevelled instead, but it cannot cleave APC because of the presence of a self-inhibitory domain. The destruction complex remains active to keep β-catenin levels low, and Wnt target genes that promote proliferation are not transcribed.

Wnt activity is highest at the base of the crypt and gradually decreases as cells differentiate and migrate to the lumen (Fig. 1a).

The 'canonical' Wnt signalling pathway hinges on the stability of a protein called β-catenin. In the absence of extracellular Wnt, β-catenin in the cell is captured by a complex of proteins that facilitates its destruction by tagging it with ubiquitin proteins. When Wnt binds to its cell-surface receptor, Frizzled, the destruction complex is inhibited, enabling β-catenin to accumulate and move into the nucleus. Here, it binds to transcription-factor proteins that mediate gene expression[5].

The authors found three variants of caspase-5 expressed in the intestine. These are generated by a post-transcriptional process called alternative splicing, which combines various protein-coding regions of a single gene to generate different messenger RNAs, resulting in distinct proteins. The longer 'isoforms', caspase-5a and caspase-5b, resemble caspase-4, whereas the shorter caspase-5c lacks the CARD and therefore does not bind to bacterial lipids to form an inflammasome.

Instead, caspase-5c has an unexpected role in augmenting Wnt signalling in transit-amplifying cells. The authors showed that, through its catalytic domain, caspase-5c binds to proteins in the Dishevelled family, which associate with the destruction complex. This enables caspase-5c to cleave the complex's main scaffolding protein, APC, causing the complex to disassemble and β-catenin to accumulate and translocate to the nucleus, where it regulates the expression of genes that promote proliferation and progenitor-cell maintenance (Fig. 1b).

All three isoforms can bind to Dishevelled proteins because they all maintain their catalytic domain, but only caspase-5c is able to cleave APC. In stem cells at the base of the crypt, high concentrations of Wnt repress *CASP5* expression, suggesting a possible negative-feedback mechanism that keeps caspase-5c low in intestinal stem cells. Caspase-5c expression is highest in transit-amplifying cells, but as cells differentiate, expression of caspase-5a and caspase-5b increases and that of caspase-5c declines sharply. Therefore, the destruction complex remains active in cells in the upper part of the crypt to reduce β-catenin levels (Fig. 1c).

Finally, the authors demonstrated a role of caspase-5c in tissue repair. Using cellular intestinal models called colonic organoids, the authors could manipulate the expression of different caspase-5c isoforms. Impressively, caspase-5c overexpression, along with lowered expression of the other isoforms, increased the size of human colonic organoids by about 50%, reflecting an increase in proliferation. The authors also found that caspase-5c expression is increased in colonic organoids that are subjected to radiation damage and in the guts of people with inflammatory bowel disease. This suggests that the augmentation of transit-amplifying-cell proliferation by caspase-5c could be crucial for repopulating damaged intestinal epithelium.

Jia and colleagues' findings open up several avenues for further investigation. There are strong hints, which the paper begins to explore, that caspase-5c has a role in colorectal cancer. The gene that encodes APC is often the first to be mutated in many forms of colorectal cancer, and these mutations cause persistent activation of Wnt signalling, driving uncontrolled proliferation. Notably, many cancer mutations truncate APC in the same N-terminal region as that cleaved by caspase-5c. Inhibiting caspase-5c might therefore be an effective way to dampen Wnt signalling to treat cancer.

To probe the function of caspase-5c in other tissues where Wnt and β-catenin signalling are important, researchers could generate a genetically modified 'knock-in' mouse that encodes *CASP5* and its regulatory regions. Indeed, a knock-in mouse that expresses human *CASP4* instead of mouse *Casp11* has been invaluable for understanding why humans are so vulnerable to sepsis[4]. This approach could also be useful for testing whether caspase-5c contributes to recovery from other types of gut injury, including infections and autoimmune conditions such as coeliac disease, graft-versus-host disease and complications of cancer immunotherapies.

Caspase-5c's protease activity seems to be required for tissue repair, but it is not yet clear how it can be activated without a CARD, which facilitates the formation of caspase pairs (dimers) and their activation through self-cleavage[6]. One possibility is that

caspase-5c is activated when it is recruited to the β-catenin destruction complex, perhaps as a dimer or other multimer. Resolving the structure of caspase-5c bound to the destruction complex could help to clarify this, and also reveal how caspase-5c cleaves APC.

Gasdermin D and pro-IL-18 (the precursor to the inflammatory protein interleukin-18; IL-18) are expressed throughout the intestinal epithelium and are substrates of the full-length caspase-5 (ref. 4). Jia *et al.* show that caspase-5c cleaves gasdermin D more efficiently than do the other isoforms, but they did not examine whether gasdermin D and pro-IL-18 are cleaved by caspase-5c in transit-amplifying cells in injured intestines. Because IL-18 promotes intestinal repair, and cells expressing the IL-18 receptor are found in the base of the crypts[7], it might be that the processing of gasdermin D and pro-IL-18 by caspase-5c, and the secretion of IL-18, play a part in intestinal repair.

Although there was no evidence of cell death in the proliferating transit-amplifying cells that express caspase-5c, it is still possible that caspase-5c activates gasdermin D (and perhaps other gasdermins) to form pores and cause local inflammation. Some cell types can survive the formation of gasdermin pores because the membrane-repair machinery removes them. The surviving cells are said to be 'hyperactivated' because, instead of undergoing cell death, they continue to release large amounts of inflammatory mediators[8]. In fact, gasdermin-D-mediated release of metabolite molecules from hyperactivated immune cells boosts muscle regeneration by increasing the proliferation of muscle stem cells[9]. If transit-amplifying cells are hyperactivated after intestinal injury, it would be worth characterizing their 'secretome' — all of the proteins, metabolites and other small molecules they secrete — and the effects of these molecules on intestinal repair.

Jia and colleagues' work aligns with mounting evidence that cell-death mediators have a broader role in tissue repair. For example, caspases that are involved in apoptosis — a form of controlled cell death — drive the proliferation needed for tissue regeneration in invertebrates such as fruit flies (*Drosophila melanogaster*) and *Hydra*[10]. In mammals, apoptosis-activating caspases promote liver regeneration, wound healing and proliferation in non-apoptotic cells close to those undergoing apoptosis[10]. They also promote abnormal tissue growth that precedes cancer[10]. In addition to the role of gasdermin D in muscle regeneration, gasdermins B and C, which are expressed in the intestinal and respiratory tracts, have been linked to tissue repair[11,12]. A better understanding of how caspase-5c and the gasdermins contribute to tissue repair could lead to therapeutic strategies for treating diseases that damage the intestine.

Ayijiang Yisimayi and **Judy Lieberman** are in the Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts 02115, USA, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA.
e-mail: judy.lieberman@childrens.harvard.edu

1. Jia, B. et al. *Nature* **652**, 1362–1374 (2026).
2. Wang, C. *et al. Sci. Adv.* **12**, eadz4623 (2026).
3. Bai, Y., Pan, Y. & Liu, X. *Nature Rev. Mol. Cell Biol.* **26**, 501–521 (2025).
4. Elkayam, E., Gervais, F. G., Wu, H., Crackower, M. A. & Lieberman, J. *Nature Rev. Immunol.* **25**, 558–568 (2025).
5. Kretzschmar, K. & Clevers, H. *Dev. Biol.* **428**, 273–282 (2017).
6. Renatus, M., Stennicke, H. R., Scott, F. L., Liddington, R. C. & Salvesen, G. S. *Proc. Natl Acad. Sci. USA* **98**, 14250–14255 (2001).
7. Winsor, N. J. *et al. Proc. Natl Acad. Sci. USA* **122**, e2417149122 (2025).
8. Zanoni, I. *et al. Science* **352**, 1232–1236 (2016).
9. Chi, Z. *et al. Nature* **634**, 1168–1177 (2024).
10. Fogarty, C. E. & Bergmann, A. *Cell Death Differ.* **24**, 1390–1400 (2017).
11. Keller, A. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2025.09.24.678435 (2025).
12. Rana, N. *et al. Cell* **187**, 1011–1015 (2024).

The authors declare no competing interests.
This article was published online on 22 April 2026.