Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multimodal Gene Expression Atlas Resolves the Cellular Context of Heterosis in Maize Roots

Peng Yu

pengyu.yu@tum.de

Technicial University of Munich https://orcid.org/0000-0003-1670-8428

Danning Wang

Plant Genetics, TUM School of Life Sciences, Technical University of Munich (TUM), Freising 85354, Germany

Can Wang

Shenzhen Digital Cloud Valley Biotechnology Co., Ltd, Shenzhen, China

Yaping Zhou

Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53117, Germany

Guoliang Wang

Beijing Agro-Biotechnology Research Center

Marion Pitz

Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53117, Germany

Jing Wang

BGI Tech Solutions, BGI Genomics, Wuhan, China

Yiji Xia

Hong Kong Baptist University Department of Biology https://orcid.org/0000-0002-9543-8341

Sanjie Jiang

Department of Biology, Hong Kong Baptist University, Kowloon 999077, Hong Kong, China

Xinping Chen

Southwest University https://orcid.org/0000-0002-6245-0133

Frank Hochholdinger

Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53117, Germany

Biological Sciences - Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11761315/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

A Multimodal Gene Expression Atlas Resolves the Cellular Context of Heterosis in Maize Roots

Danning Wang[1,2,4,7,#], Can Wang[5,#], Yaping Zhou[2,#], Guoliang Wang[1,9,#], Marion Pitz[2], Jing Wang[8], Yiji Xia[6,10,11], Sanjie Jiang[6,*], Xinping Chen[3,*], Frank Hochholdinger[2,*], Peng Yu[1,4*]

[1] Plant Genetics, TUM School of Life Sciences, Technical University of Munich (TUM), Freising 85354, Germany

[2] Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53117, Germany

[3] College of Resources and Environment, and Academy of Agricultural Sciences, Southwest University, Chongqing 400715, China

[4] Emmy Noether Group Root Functional Biology, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53117, Germany

[5] Shenzhen Digital Cloud Valley Biotechnology Co., Ltd, Shenzhen, China

[6] Department of Biology, Hong Kong Baptist University, Kowloon 999077, Hong Kong, China

[7] Plant Breeding, TUM School of Life Sciences, Technical University of Munich, Freising 85354, Germany

[8] BGI Tech Solutions, BGI Genomics, Wuhan, China

[9] Institute of Biotechnology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

[10] College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo, 315100, China;

[11] State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong SAR, China

[#] These authors equally contributed to this work.

[*] To whom correspondence should be addressed:

Frank Hochholdinger: hochholdinger@uni-bonn.de

Xinping Chen: chenxp2017@swu.edu.cn

Sanjie Jiang: jiangsanjie@hkbu.edu.hk

Peng Yu: pengyu.yu@tum.de (lead contact)

Abstract

Cross-fertilized hybrid crops outperform their genetically distinct parents, a phenomenon called heterosis. Despite the extensive use of heterosis in crop breeding for almost a century, the mechanism of how hybrid gene regulation relates to increased cellular growth remains poorly understood. Here we integrate single-cell RNA sequencing, spatial transcriptomics, laser-capture microdissection and quantitative anatomy to define the cellular organization of heterosis in maize (*Zea mays*) roots. An interactive atlas enables exploration of non-additive genes across these datasets (http://93.93.112.125/). Across nine hybrid–parent combinations, non-additive expression varied among root cell types and developmental states, with the strongest responses in epidermal and cortex populations. Although the genes involved differed substantially among hybrids, the extent of cortex non-additivity was positively associated with cortex cell-size heterosis. Developmental and spatial analyses revealed non-additive expression to cortex differentiation and radial position, while revealing hybrid-specific variation in these relationships. Computational virtual knockout analyses implicated carbon–nitrogen metabolic networks may drive the differences of heterosis between parental combinations, and further genomic mapping together with spatial expression nominated *THP3/GOT1*, a gene involved in nitrogen assimilation, as a candidate associated with root heterosis. These findings establish developmental and cellular context as organizing dimensions of heterosis and provide a framework for understanding how distinct hybrid regulatory programmes promote root growth.

Key words

Heterosis, Maize, Root, Single-cell sequencing, Spatial transcriptomics

Main

Heterosis, the superior performance of cross-fertilized hybrids relative to their genetically distinct parents, has driven productivity gains across many crops and transformed modern agriculture (Duvick, 2001). Maize (*Zea mays* L.) is a central model for investigating its genetic and physiological basis (Birchler et al., 2010), yet after more than a century of research, how parental genetic divergence is translated into enhanced growth in hybrids remains unresolved (Schnable and Springer, 2013). A key challenge is to connect molecular differences between hybrids and their parents to the cellular processes that build complex organs. This requires resolving hybrid-associated regulation in anatomical space, following its dynamics during development and relating it to growth in the corresponding cell populations.

Roots offer a tractable system for addressing these questions. Root architecture and anatomy determine the capacity of plants to acquire water and nutrients and to adapt to heterogeneous soils (Yu et al., 2016; Lynch, 2022). Maize root systems were substantially reshaped during domestication and breeding (Yu et al., 2024), and heterosis emerges early in root development, affecting root length, thickness and branching (Hochholdinger and Yu, 2025). These organ-level phenotypes arise from the coordinated proliferation, expansion and differentiation of distinct root cell populations, yet the contribution of individual cell states to root heterosis remains poorly resolved.

Roots are organized along intersecting longitudinal and radial axes. Longitudinally, cells progress from division through elongation to terminal differentiation (Rogers and Benfey, 2015). Radially, epidermal, cortical, endodermal and vascular tissues occupy distinct positions and perform specialized functions (Tian et al., 2014). Lateral roots and root hairs further diversify cellular activities and developmental outcomes within the organ (Orman-Ligeza et al., 2013; Dolan, 2017). Hybrid–parent differences could therefore arise from changes in the abundance of particular cell states, altered developmental progression or transcriptional regulation within otherwise comparable states. Discriminating among these possibilities is essential for understanding how heterosis is expressed within a complex organ.

Bulk transcriptomic studies have identified widespread non-additive expression that is different from the parental mean (Paschold et al., 2012; Baldauf et al., 2018). Tissue-averaged measurements, however, cannot distinguish regulatory changes within cells from shifts in the cellular composition of the sampled tissue (Gouran and Brady, 2024). Single-cell and spatial transcriptomics overcome these limitations by resolving cellular identity, developmental state and anatomical context (Minne et al., 2022; Denyer and Timmermans, 2022). In plants, these methods have illuminated ground-tissue patterning, phloem maturation, lateral-root initiation and the regulation of cell elongation, while enabling comparisons of cellular programmes across species (Kajala et al., 2021; Ortiz-Ramírez et al., 2021; Roszak et al., 2021; Omary et al., 2022; Guillotin et al., 2023; Nolan et al., 2023). However, these approaches have largely been used to characterize individual genotypes, developmental programmes or stress responses, and have only rarely been applied to systematic hybrid–parent comparisons. Consequently, it remains unknown whether non-additive regulation is concentrated in particular cell types, developmental states or anatomical positions, and whether genetically distinct hybrids converge on common cellular contexts of heterosis despite involving different genes.

A multimodal atlas defines the cellular framework of maize root heterosis

To determine where hybrid-associated transcriptional differences arise, we first established an anatomically anchored reference atlas of maize root cell types by integrating laser-capture microdissection (LCM), spatial transcriptomics and single-cell RNA sequencing (Supplementary Fig. S1). LCM profiling resolved 11 root tissues or cell populations, including the epidermis, outer and inner cortex, endodermis, phloem-pole and xylem-pole pericycle, proto-phloem, proto-xylem, metaxylem, cells between metaxylem poles and pith. Multidimensional scaling separated samples according to anatomical identity, with their transcriptional relationships closely recapitulating the radial organization of the root from inner vascular domains to the outer cortex and epidermis (Fig. 1a). For each population, genes differentially expressed relative to the remaining LCM samples were ranked by log counts per million, and up to 50 highly expressed candidates were retained as reference markers (Supplementary Fig. S2, Table S1). This strategy generated molecular reference profiles for root populations that are difficult to resolve by anatomy alone, including the outer, middle and inner cortex layers, xylem-associated cells and phloem-pole pericycle, and established an integrated framework for cross-platform annotation.

The spatial transcriptomic dataset comprised 34 root sections of B73, A554 and B73 × A554, yielding 34,463 spatial spots at 25 μm × 25 μm resolution and 16,383 detected genes after filtering lowly expressed genes and low-quality cells (Supplementary Fig. S3; Table S2). Spatial clustering constrained by anatomical position, tissue-enriched expression and LCM reference profiles, resolved the major radial domains of the root, including the epidermis, outer, middle and inner cortex, endodermis, pericycle, vascular tissues, metaxylem-associated cells and pith (Fig. 1b, Supplementary Fig.S4, Table S3). Projection of these assignments onto a representative transverse root section recapitulated the expected radial organization (Fig. 1c, Supplementary Fig.S5). Thus, integration of LCM-derived marker profiles with spatial transcriptomic coordinates generated a transcriptome-resolved and anatomically precise map of the maize root.

We next profiled 21 primary-root samples collected five days after germination, comprising three biological replicates of B73, nine paternal inbred lines representing non-stiff-stalk, stiff-stalk and tropical genetic groups, and the corresponding nine B73-derived F_1-hybrids (Supplementary Fig. S6). After quality control, 91,157 high-quality cells were retained (Supplementary Fig. S7; Table S4) and integrated into 15 transcriptional clusters (Table S5). Cell identities were assigned by jointly considering cluster-enriched expression, LCM reference profiles, spatial localization and published marker genes (Fig. 1d, Table S6).

The marker-expression matrix revealed coherent cluster-enriched programmes spanning epidermal, cortical, vascular and initial-cell domains (Fig. 1d). The integrated single-cell embedding resolved the corresponding cell populations, including epidermis, cortex, endodermis, pericycle, phloem, xylem, stele and pith, while preserving their developmental and anatomical relationships (Fig. 1e). Marker genes localized to defined anatomical domains in the spatial sections (Fig. 1f) were selectively enriched in the corresponding single-cell clusters (Fig. 1g) and showed concordant expression in the LCM-derived reference profiles (Fig. 1h). The reciprocal agreement between spatial localization and single-cell

enrichment demonstrated that the same marker programmes could be mapped across independent platforms and assigned to their corresponding anatomical populations.

Pseudobulk principal-component analysis provided a sample-level assessment of transcriptional relationships and reproducibility (Supplementary Fig. S8). B73 replicates clustered closely, whereas paternal inbred lines occupied distinct regions according to their genetic group. Most hybrids were positioned between B73 and their corresponding paternal parent, but their displacement within this parental space varied among crosses. These patterns indicated that hybrid roots did not converge on a single uniform transcriptional state but instead retained genetic-background-dependent transcriptional relationships. Although bulk transcriptomic studies have previously identified widespread non-additive and complementary expression in maize hybrids (Hoecker et al., 2008), organ- or tissue-level measurements cannot distinguish regulatory changes within cells from shifts in the abundance of cell populations. By linking LCM-derived molecular references to spatial coordinates and single-cell profiles, our atlas defined comparable anatomical and developmental units across genotypes. This framework therefore enabled us to determine whether non-additive expression is distributed broadly across the root or concentrated in specific cell states.

Cell identity and radial position organize hybrid non-additivity

With this cross-platform atlas providing matched cellular, developmental and anatomical units across genotypes, we next asked where non-additive expression emerges in the hybrid root. We defined non-additively expressed genes (NAGs) by comparing hybrid expression with the mean expression of the two corresponding parents within each matched cell population (Paschold et al., 2012). Analyses were restricted to genes expressed in at least 10% of cells per cell population (see Methods; Table S7). The fraction of expressed genes classified as non-additive varied markedly among cell populations. It was highest in mature cortex (Cortex I) and epidermal cells, whereas initials, columella and several internal tissues showed lower levels (Fig. 2a). Independent hybrid–parent comparisons reproduced the recurrent prominence of cortex and epidermal populations, although the magnitude of non-additivity varied among crosses (Supplementary Fig. S9).

Parental genetic background further influenced both the magnitude and cellular distribution of NAGs. Crosses between B73 and the non-stiff-stalk lines A554, Oh40B and Oh43 showed the strongest and most widespread responses across cell populations (Supplementary Fig. S9). Tropical crosses also accumulated NAGs across multiple populations, but their responses were more heterogeneous; B73 × NC352 displayed a comparatively broad non-additive profile. By contrast, the stiff-stalk × stiff-stalk combinations B73 × A680 and B73 × B37 showed the lowest and most restricted NAG proportions. Thus, genetic background influenced not only how much non-additive expression occurred, but also the cellular contexts in which it was deployed.

NAG-frequency and cell-type specificity were related but distinct properties. Most NAGs were detected in more than one population, and only 1–10% were restricted to a single population (Fig. 2b). Initials and columella contained the largest relative fractions of population-specific NAGs despite their comparatively low overall levels of non-additive expression. Non-additive transcription therefore comprised a broadly shared programme together with a smaller, population-restricted component.

NAG recurrence and regulatory mode showed a pronounced radial organization. From the outer epidermal and cortical populations towards the inner root tissues, the fraction of NAGs detected in only one hybrid combination increased, indicating progressively greater hybrid specificity (Fig. 2c). In parallel, underdominant expression became more prevalent towards inner populations, whereas outer tissues contained a larger recurrent component of NAGs (Fig. 2d). Root cell identity and spatial organization was therefore associated not only with the frequency of non-additive expression, but also with its cross-hybrid reproducibility and regulatory direction. Outer tissues preferentially engaged recurrent non-additive programmes, whereas inner populations displayed more hybrid-specific and underdominant regulatory states.

To determine whether this cellular organization of non-additive expression was preserved in intact roots, we analyzed all spatially resolved observations from B73, A554 and B73 × A554 root sections and identified NAGs within anatomically defined populations (Table S8). Spatial transcriptomics independently reproduced the enrichment of non-additive expression in cortex-associated domains and revealed a relatively large population-specific component in the epidermis (Supplementary Fig. S10). Although fewer NAGs were detected spatially than by single-cell profiling, most overlapped with those identified in the corresponding single-cell populations, with the highest concordance in cortex and epidermis (Supplementary Fig. S11). This cross-platform agreement indicates that the cellular distribution of non-additivity reflects a reproducible biological organization rather than a feature of a single profiling method. Together, these analyses showed that non-additive expressions are organized by cell identity, radial position and parental genetic background, rather than distributed uniformly across the hybrid root.

Non-additive programmes accompany cellular size heterosis and metabolic specialization

Having established where non-additive regulation occurs, we next asked whether its cellular distribution is associated with anatomical heterosis and the functional programmes expressed in hybrid roots. Using the same hybrid–parent framework, we quantified cell size in the epidermis, three cortex layers, endodermis, metaxylem and pith, together with cortex and stele areas, from root-section images (Supplementary Fig. S12). Mid-parent heterosis (MPH) was calculated for each anatomical trait and hybrid (Fig. 2e; Supplementary Fig. S13; Table S9). The strongest and most extensive positive anatomical heterosis occurred in B73 × Oh40B and B73 × Oh43. Heterosis was more variable in tropical crosses, and was weakest in stiff-stalk derived hybrids: B73 × A680, B73 × B37 and B73 × NC328. Across hybrids, mean cell-size MPH was highest in metaxylem (27%), followed by inner cortex (19%) and middle cortex (18%), whereas epidermal cell size showed a slightly negative mean MPH (−1.5%) and endodermal cell size did not significantly deviate from the mid-parent expectation (Supplementary Fig. S13). These comparisons showed that the frequency of NAGs and the magnitude of anatomical heterosis were related but not equivalent: populations with frequent non-additive expression were not necessarily those with the largest mean increase in cell size.

Within individual populations, however, the extent of transcriptional non-additivity was positively associated with cellular heterosis across the nine hybrids (Fig. 2f). NAG abundance was positively correlated with the MPH of the corresponding cell-size trait in the middle cortex (r = 0.75, P = 0.019),

epidermis (r = 0.68, P = 0.045), inner cortex (r = 0.71, P = 0.032) and metaxylem (r = 0.72, P = 0.029). The association was weaker in the outer cortex (r = 0.60, P = 0.09), pith (r = 0.65, P = 0.059) and absent in endodermal populations (r = 0.32, P = 0.40; Supplementary Fig. S14). Thus, greater cell-state-resolved transcriptional reprogramming tended to accompany stronger cellular heterosis across several anatomically distinct populations, although NAG abundance alone did not determine cell size. These associations are correlational and do not establish a causal relationship.

Functional analysis of recurrent NAGs revealed convergence at the level of biological programmes despite limited recurrence of individual genes. Cortex NAGs were enriched for water transport, developmental regulation and hormone- and stress-responsive processes, whereas epidermal NAGs were associated with water transport, cell-wall organization, stress adaptation, translational capacity and interactions with the root environment (Supplementary Figs. S15 and S16).

We next compared the functional composition of genes preferentially expressed in hybrid and parental cells. Across most root populations, parent-enriched genes contained a higher proportion of response-to-stimulus Gene Ontology (GO) terms than hybrid-enriched genes (Fig. 2h). In Cortex I mature, response-to-stimulus terms accounted for 31.2% of the assigned GO level-3 categories in the parent-enriched set, compared with 11.3% in the hybrid-enriched set. Conversely, metabolic-process terms represented 44.1% of categories in the hybrid-enriched set, compared with 27.1% in the parent-enriched set (Fig. 2g). Across root populations, this shift towards metabolic functions was most pronounced in selected differentiated cortex states, although its magnitude varied among tissues (Fig. 2i). Parental cells therefore retained stronger stimulus- and stress-responsive transcriptional signatures, whereas hybrid cells preferentially engaged metabolic programmes within shared cellular contexts. This functional divergence provides a potential explanation for how distinct hybrid combinations can achieve enhanced anatomical performance without uniformly increasing transcription across the root.

Finally, to examine the transcriptional basis of the contrasting anatomical responses of B73 × Oh40B and B73 × A554, we intersected cortex NAGs from B73 × Oh40B with genes differentially expressed between the two hybrids. The intersecting genes were enriched for metabolic processes, responses to chemical stimuli, cell–cell communication and stress-related functions (Supplementary Fig. S17). The corresponding epidermal comparison identified metabolic and stress-response terms together with cell-wall, transport and environmental-response functions (Supplementary Fig. S18). Thus, shared cellular contexts can contain distinct hybrid-specific functional programmes, providing a plausible basis for the convergence of anatomical phenotypes despite limited recurrence of individual genes.

Developmental and radial coordinates organize cortex heterosis

Because the cortex emerged as a major cellular context of non-additive expression, we next asked whether cortex heterosis is associated with developmental progression and radial position. We examined three previously annotated cortex transcriptional states—Cortex early, Cortex I mature and Cortex II mature (Fig. 3a). Cells from non-stiff-stalk, stiff-stalk and tropical-line materials overlapped within the shared embedding, but their relative occupancy differed among genetic groups. In particular, hybrid cortex cells from tropical-line and stiff-stalk crosses were significantly closer to their non-B73 parental profiles than to B73, whereas the non-stiff-stalk hybrids showed no significant preference for

either parent (Fig. 3b; Supplementary Fig. S19). We next projected published maize root developmental programmes onto this cortex atlas (Ortiz-Ramírez et al., 2021) and reconstructed cortex development with Monocle3 (Cao et al., 2019). The meristematic score (M_Score) was highest in earlier states, whereas the elongation-associated score (E_Score) increased towards later states (Fig. 3c; Supplementary Fig. S20). Consistent with these programme-level patterns, Monocle3 reconstructed a continuous pseudotime trajectory spanning early to mature cortex states (Fig. 3d). Thus, cortex cells from distinct genetic backgrounds occupied a common developmental landscape, but differed in their representation and parental resemblance within that landscape.

The NAG module score varied across the cortex embedding (Fig. 3e). Eight of the nine crosses showed positive associations between NAG scores and E_Score, together with negative associations with M_Score, whereas B73 × Oh43 showed the opposite orientation (Fig. 3f, Table S10). Non-additive expression was therefore coupled to cortex maturation rather than distributed uniformly along the developmental trajectory, although the strength and direction of this coupling varied among hybrids. In the B73 × Oh40B hybrid, the score for NAGs exhibiting positive expression MPH was strongly associated with E_Score (Pearson's r = 0.774, P < 0.001; Fig. 3g), whereas negative-MPH NAGs showed no comparable association (Supplementary Fig. S20d).

Radial position provided a second coordinate for cortex heterosis. Spatial annotation separated the outer, middle and inner cortex, allowing the positive-MPH NAG score to be mapped within the cortex while excluding other tissues (Fig. 3h, i; Supplementary Fig. S21a-b). The positive-MPH NAG score was differentially distributed across cortex layers, and section-level medians differed significantly in paired comparisons after Benjamini–Hochberg correction (Fig. 3j). Spatial profiles of negative-MPH NAGs provided complementary radial comparisons (Supplementary Fig. S21c-d). The three-layer organization was reproducibly observed across additional sections, indicating that the radial pattern was not attributable to a single root section. Together, these results define cortex heterosis in two complementary spatial dimensions: a hybrid-dependent longitudinal component associated with developmental maturation, and a reproducible radial component distributed across cortex layers.

These findings add developmental and spatial dimensions to classical dominance, overdominance and complementation models, which describe parent–hybrid relationships but do not specify where or when these regulatory patterns emerge. Within individual roots, non-additive expression varied with developmental state and radial position, while its relationship with maturation differed among parental combinations. The cross-specific orientation of these relationships is consistent with previous evidence for single-parent complementation and hybrid-specific regulatory variation in maize (Baldauf *et al.*, 2022; Janko *et al.*, 2024; Pitz *et al.*, 2025), suggesting that similar anatomical outcomes may arise through distinct regulatory architectures.

Carbon–nitrogen network analysis prioritizes *THP3/GOT1* as a context-dependent candidate

Having established that non-additive expression is concentrated in differentiated cortex states and organized along developmental and radial axes, we next asked whether these cell-state-specific programmes converge on regulatory networks that could connect parental transcriptional divergence to heterotic cell expansion. We focused on B73 × A554 and B73 × Oh40B because both are B73 × non-

stiff-stalk crosses with strong non-additive expression and positive anatomical heterosis in cortex and epidermis, but differed in the magnitude and consistency of their tissue-specific responses (Fig. 2e). This contrast provided a framework for distinguishing metabolic features shared between the two hybrids from regulatory responses that were cross-specific. Candidate genes were selected by integrating cell-type-enriched expression, NAG status, functional annotation and spatial localization (Table S11). Although these observational data supported candidate gene prioritization, they did not establish their functional roles in heterosis. We therefore performed single-gene virtual perturbations using scTenifoldKnk (Osorio *et al.*, 2022). Candidates from five functional modules were evaluated across cortex and epidermal states in both hybrids. The analysis tested whether perturbation preferentially affected NAGs, whether predicted effects differed between cortex and epidermis, and whether the affected gene sets were shared or hybrid-specific. These analyses were interpreted as predictions of regulatory-network sensitivity rather than as substitutes for experimental gene disruption.

Carbon–nitrogen, lipid and energy-metabolism genes produced the clearest cross-dependent separation. Among the top 10% of genes ranked by predicted perturbation, NAG enrichment showed opposite epidermis-versus-cortex biases in the two hybrids (Fig. 4a; Supplementary Fig. S22a). This cross-specific partitioning was consistent with the anatomical measurements: B73 × Oh40B showed broad positive heterosis across cortex-associated traits, whereas B73 × A554 displayed more heterogeneous tissue-level responses (Fig. 2e). Thus, the same broad metabolic category was predicted to engage different cellular programmes in the two hybrid backgrounds. We next examined whether perturbations affected common or distinct sets of NAGs (Fig. 4b). In B73 × A554, perturbation signatures were heterogeneous, with limited or uneven overlap among candidates. By contrast, in B73 × Oh40B, the perturbation signature of *Zm00001d043382*, which encodes glutamate–oxaloacetate transaminase 1 (GOT1) and corresponds to *TEOSINTE HIGH PROTEIN 3* (*THP3*) (Huang *et al.*, 2026a), substantially overlapped with the signatures of four other candidate genes and showed greater similarity to their predicted responses. This recurrent effect across otherwise distinct candidates prioritizes *THP3/GOT1* as a potential regulatory anchor within the Oh40B-associated carbon–nitrogen network, while its more limited correspondence in B73 × A554 indicates that its network effects are hybrid-dependent. Gene Set Enrichment Analysis (GSEA) further showed preferential enrichment of positive-MPH NAGs among genes receiving stronger predicted perturbation effects, whereas negative-MPH NAGs showed the opposite trend (Supplementary Fig. S22b). In B73 × Oh40B, perturbed genes were enriched for root-hair elongation, meristem growth, auxin signaling, morphogenesis, cell-wall organization, systemic acquired resistance and cytoskeletal functions (Fig. 4c). These predicted responses connect metabolic perturbation with developmental and structural processes relevant to root-cell expansion.

Using spatially validated marker genes, we delineated epidermal and cortical domains along longitudinal root sections using in situ hybridization, thereby providing anatomically defined regions for tissue-resolved expression analysis (Supplementary Fig. S23, Table S12). Within these marker-defined regions, the eight-gene module showed strong tissue- and cross-specific expression MPH (Fig. 4d; Supplementary Fig. S24). In the epidermis of B73 × Oh40B, *SACD9* (stearoyl-acyl-carrier-protein desaturase 9), *PDC3* (pyruvate decarboxylase 3) and *THP3/GOT1* reached +115%, +73% and +69%

MPH, respectively, compared with +47%, +41% and +19% in B73 × A554. The spatial measurements therefore provided an independent validation that the candidates identified by virtual perturbation were regulated in a tissue- and hybrid-dependent manner.

QTL mapping based on IBM-RIL-derived population for MPH of root-tip number, total root length and total root volume identified a shared signal near the distal region of chromosome 3 (Fig. 4e; Table S13). Because *THP3/GOT1* was located near this shared QTL interval, it provided a genetically supported candidate entry point into the carbon–nitrogen network. To test whether this association extended beyond the original hybrid panel, we analyzed an IBM-RIL-derived population crossed to B73 and Mo17. At the *THP3/GOT1* locus, B73/Mo17 heterozygous genotypes showed higher MPH for total root length, root-tip number and total root volume than B73/B73 homozygous genotypes (Fig. 4f). This independent population-level association supports a relationship between heterozygosity at the candidate region and enhanced root performance, while remaining insufficient to establish molecular causality.

Finally, longitudinal spatial maps generated on a shared expression scale resolved genotype- and cross-dependent *THP3/GOT1* expression in B73, Oh40B, A554 and their corresponding hybrids (Fig. 4g). Expression MPH was consistently higher in the epidermis than in the cortex in both crosses, reaching 69% versus 19% in B73 × Oh40B and 19% versus 12% in B73 × A554 (Fig. 4h). Together, the perturbation, spatial-expression and QTL-based analyses converge on *THP3/GOT1* as a tissue-restricted, hybrid-dependent candidate linking carbon–nitrogen-associated regulation with maize root heterosis.

Discussion

Since the first description of hybrid maize by Shull (1908) and East (1908), different genetic models have been suggested to explain heterosis (reviewed in Hochholdinger and Yu, 2025). These models explain how parental alleles may interact, within a single gene or across different genes, to produce hybrid vigor, but they do not explain where these interactions are expressed or why a single hybrid can exhibit different modes of inheritance for different traits. Subsequent genomic studies established that heterosis is polygenic, network-based and dependent on parental divergence rather than on a universal heterotic gene (Birchler *et al.*, 2003; Lippman and Zamir, 2007; Birchler *et al.*, 2010; Schnable and Springer, 2013). Our findings add a cellular and spatial dimension to this framework: heterosis is not a uniform molecular property of the hybrid genome, but a conditional property of defined developmental and anatomical cell states.

Root studies in maize, wheat and rice have shown that heterosis emerges early and affects primary-root growth, branching and cellular expansion (Hoecker *et al.*, 2006; Wang *et al.*, 2006; Zhai *et al.*, 2013). Tissue-resolved analyses further demonstrated that non-additive and single-parent- expression complementation varies among meristematic, elongation, cortex and stele compartments (Paschold *et al.*, 2012, 2014; Baldauf *et al.*, 2016, 2018, 2022; Zhou *et al.*, 2019). Our work advances this view from broadly dissected tissues to cell states and radial positions. It suggests that differentiation creates a developmental competence window in which parental regulatory differences become translated into non-additive transcription and cellular growth. Thus, dominance, complementation and overdominance may represent local outcomes of developmental state and tissue position rather than fixed properties of a hybrid genotype.

The limited recurrence of individual non-additive genes across hybrids is therefore compatible with a shared mechanism. Different parental combinations may use distinct genes to converge on related programmes for cell expansion, transport, development and metabolism, a many-to-one relationship between regulatory variation and cellular phenotype. This interpretation integrates previous evidence for single-parent complementation and cross-specific regulatory architectures (Baldauf *et al.*, 2018, 2022; Janko *et al.*, 2024; Pitz *et al.*, 2025) with recent chromatin- and eQTL-based models in which regulatory hubs and trans-eQTL hotspots coordinate non-additive expression (Huang *et al.*, 2026b; Xu *et al.*, 2026). Chromatin-level cell studies provide potential upstream mechanisms (Jiang *et al.*, 2026), whereas our study identifies the developmental and spatial contexts in which such regulatory divergence is associated with root heterosis.

The carbon–nitrogen axis provides a plausible functional framework for this cellular dependence. Because 2-oxoglutarate, glutamate and aspartate connect carbon metabolism with nitrogen assimilation and biosynthesis (Appels and Haaker, 1991; Liu *et al.*, 2022), the chromosome-3 QTL and tissue-resolved expression of *THP3/GOT1* nominate this gene as a candidate metabolic interface. Natural *THP3* alleles have been associated with GOT1 activity and nitrogen assimilation in maize (Huang *et al.*, 2026a), but our data do not establish *THP3/GOT1* as a causal mediator or master regulator. Instead, they support a testable model in which metabolic interfaces amplify parental regulatory differences selectively in differentiated tissues with high biosynthetic demand.

Several limitations constrain this interpretation. All crosses shared B73, NAG classification was relative to parental means, virtual perturbations predicted network sensitivity rather than knockout phenotypes, and spatial expression did not measure metabolic flux. The relationship between chromatin regulation, trans-eQTL activity and the identified cell-state programmes also remains unresolved. Tissue-specific perturbation of *THP3/GOT1*, cell-resolved metabolite measurements, isotope tracing and experiments under contrasting nitrogen regimes will be required to test whether these cell-state-specific metabolic programmes causally contribute to root heterosis.

Main figures

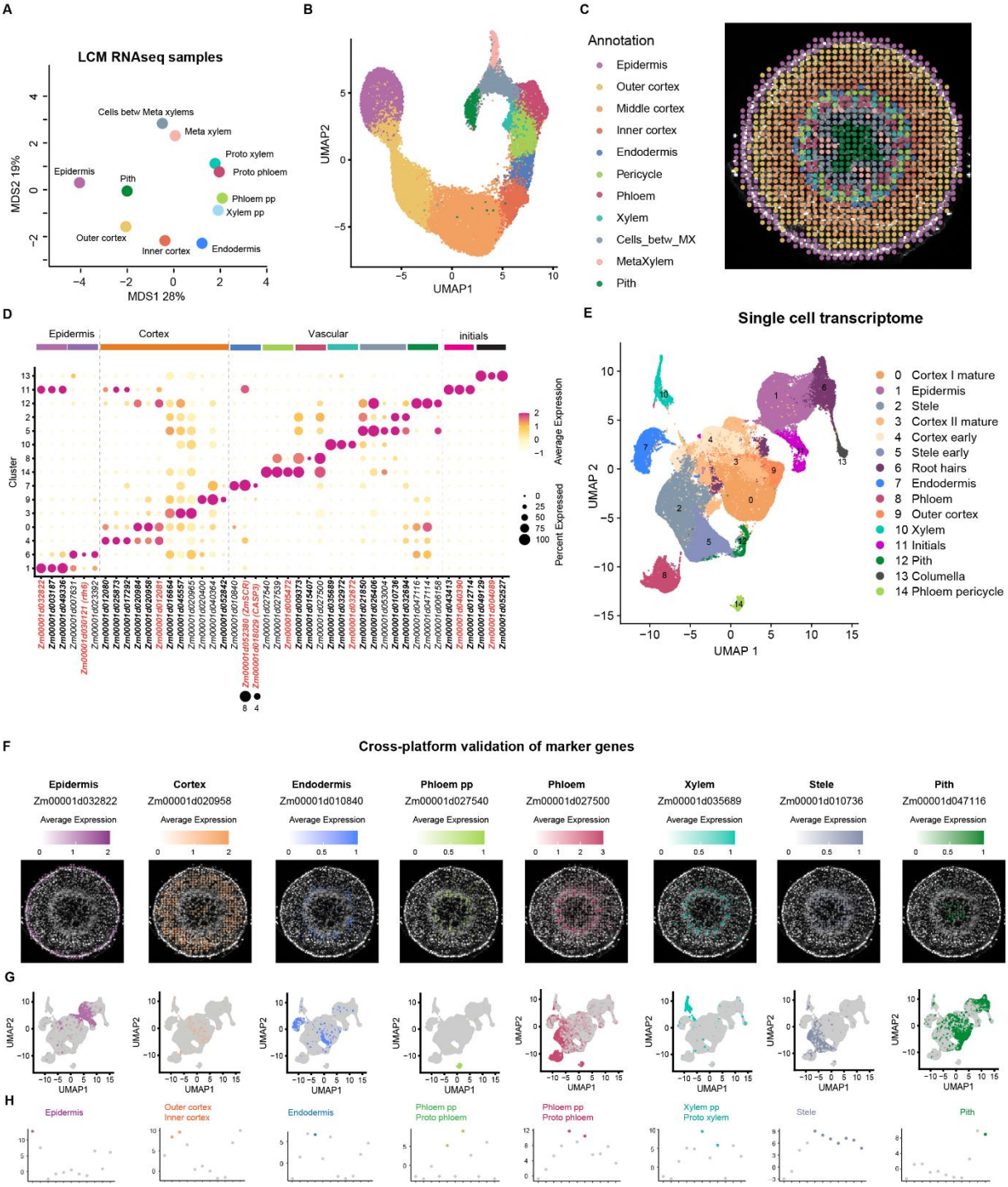


Figure 1. Multimodal integration and annotation of maize root cell types using spatial transcriptomics, laser-capture microdissection, and single-cell RNA sequencing.

(A) Multidimensional scaling (MDS) of laser-capture microdissection (LCM) RNA-seq profiles from anatomically defined maize root tissues and cell types. Each point represents one LCM-derived sample, and distances indicate transcriptomic similarity among samples. **(B)** UMAP representation of spatial transcriptomic spots, colored according to their annotated tissue identities. Spatial clusters were assigned by integrating anatomical position, tissue-enriched gene expression and LCM reference profiles. **(C)** Projection of the spatial transcriptomic annotations in (**B**) onto a representative transverse maize root section, illustrating the radial organization of the different cell types. **(D)** Dot plot showing the expression of representative marker genes across single-cell transcriptomic clusters. Different clusters

are grouped into epidermal, cortical, vascular and initial-cell domains, as indicated by the coloured bars above the plot. Dot size represents the percentage of cells expressing each gene, and dot colour indicates scaled average expression. Orange gene IDs denote markers validated by *in situ* hybridization; bold IDs indicate previously published markers; and black, non-bold IDs indicate markers newly identified using the LCM or spatial transcriptomic datasets. **(E)** UMAP visualization of the integrated single-cell RNA-seq dataset, with cells colored by their annotated cell identities. Cell-type assignments were established by jointly considering cluster-enriched marker expression, LCM reference profiles and spatial localization. **(F)** Spatial validation of representative single-cell marker genes using transverse-root spatial transcriptomic maps. Expression patterns are shown for markers of the epidermis, cortex, endodermis, phloem-pole pericycle, phloem, xylem, stele and pith. Colored points indicate spots assigned to the corresponding cell or tissue identity, confirming the anatomical localization of the single-cell clusters. (**G**) Single cell feature expression of corresponding marker genes. UMAP feature plots show the expression of marker genes corresponding to the tissue identities validated spatially in **(F)**. Cells expressing the corresponding marker genes are highlighted in cell-type-specific colors, whereas all other cells are shown in grey. (**H**) LCM-based expression of the spatially validated marker genes across anatomically defined maize root tissues. Dot plots show the expression of the corresponding markers in LCM-derived samples. Expression values were normalized as described in the Methods. Corresponding cell types are highlighted using matching colors in the scRNA-seq feature plots (G) and LCM gene-expression plots (H), with cell-type identities additionally indicated by colored labels.

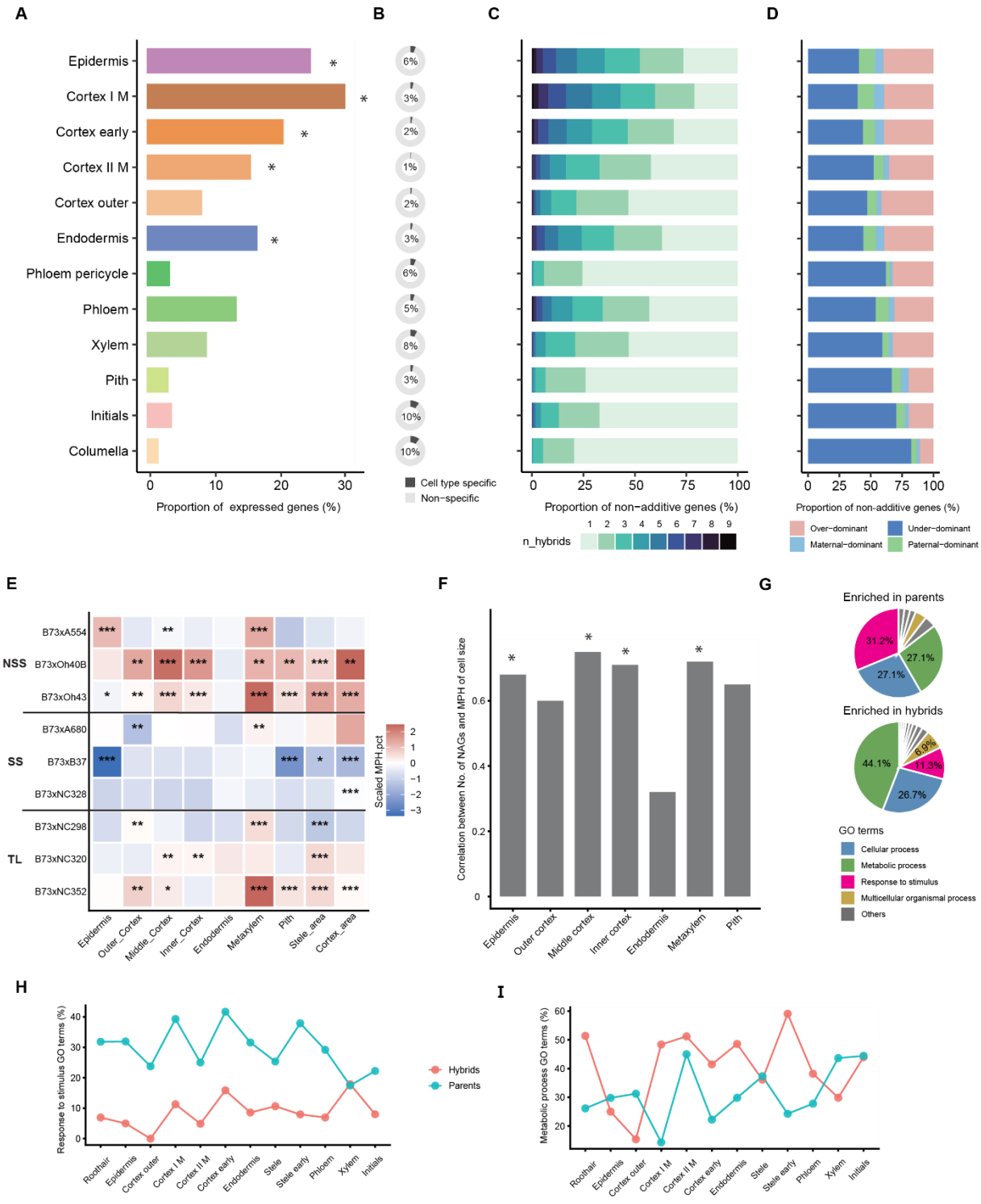
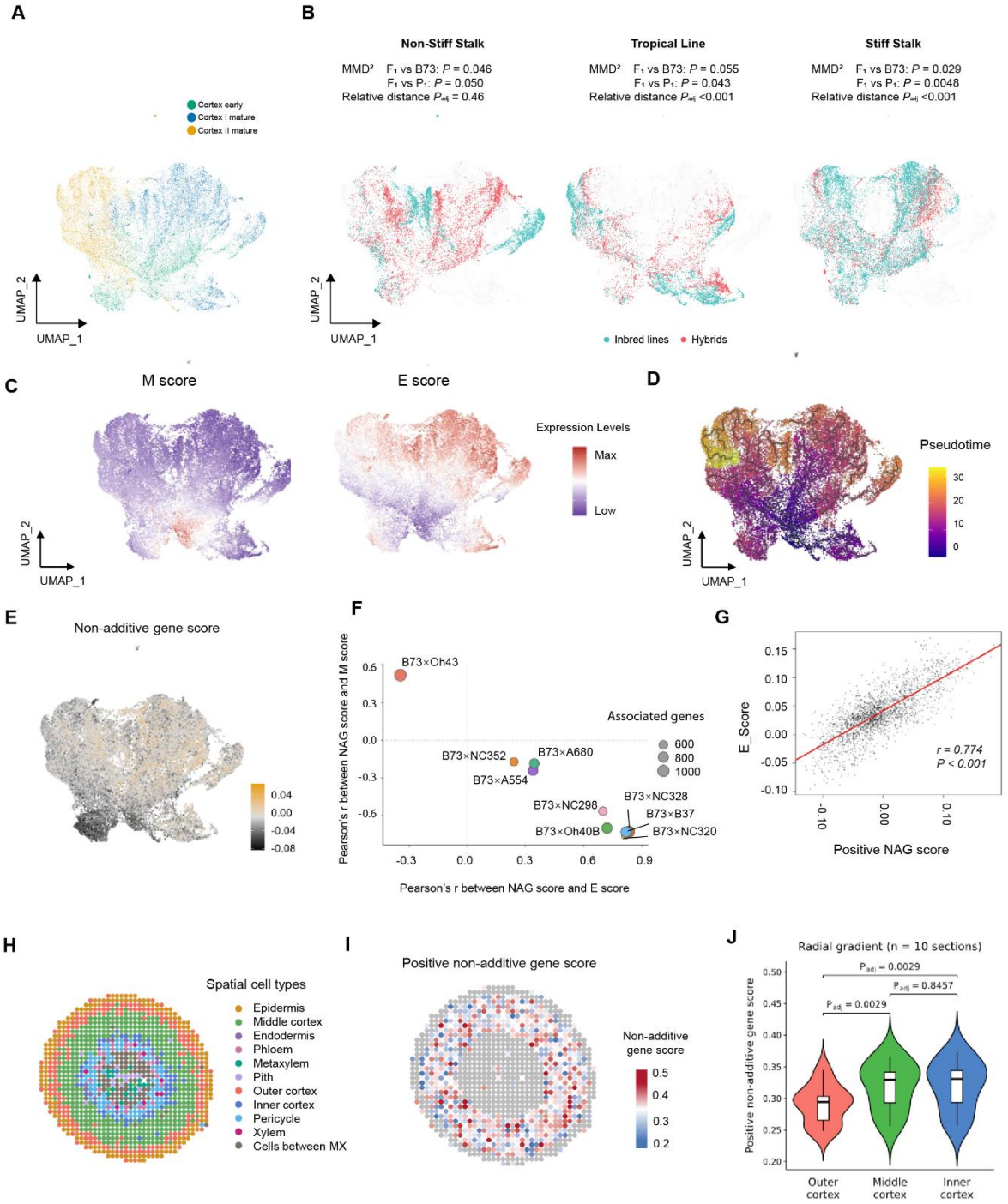


Figure 2. Cross-specific non-additive expression converges on cortex programmes linked to maize root heterosis.

(A) Percentage of expressed genes classified as non-additive across maize root cell types and developmental states. Populations are ordered according to their radial or developmental identities, with the highest proportions observed in epidermal and cortex populations. Asterisks indicate cell types in which the proportion of non-additively expressed genes (NAGs) differed significantly from the mean proportion across all other cell types (P <0.05). **(B)** Cell-type specificity of non-additively expressed genes (NAGs). Donut plots correspond to the populations shown in **A**. Black sectors indicate genes detected specifically in the indicated population, whereas grey sectors indicate genes also detected in

other populations. Values inside the donuts show the percentage of population-specific NAGs. **(C)** Recurrence of NAGs across the nine hybrid combinations. Stacked bars show the percentages of NAGs detected in one to nine hybrids for each root population. Outer populations contain a larger recurrent component, whereas inner populations contain a greater proportion of hybrid-specific NAGs. **(D)** Regulatory modes of NAGs across root populations. Stacked bars show the proportions of genes exhibiting overdominant, underdominant, maternal-dominant or paternal-dominant expression. Rows correspond to the populations shown in **A**. **(E)** Mid-parent heterosis (MPH) for root anatomical traits across nine B73-derived hybrids. Columns represent epidermal, outer-, middle- and inner-cortex, endodermal, metaxylem and pith cell-size traits, together with stele-area and cortex-area measurements. Red and blue indicate positive and negative MPH, respectively. Asterisks denote significant differences from the mid-parent expectation (*: P <0.05; **: P <0.01; ***: P <0.001). **(F)** Association between transcriptional non-additivity and cellular heterosis. Each bar represents a root cell population, and the ordinate shows the correlation across hybrids between NAG abundance and MPH for the corresponding cell-size trait. Asterisks denote significant correlations with P <0.05. **(G)** Gene Ontology level-3 functional composition of genes enriched in Cortex I mature cells of hybrids and parental inbred lines. Pie charts show the relative contributions of cellular processes, metabolic processes, responses to stimuli, multicellular-organismal processes and other categories. Percentages indicate the contribution of each category to the assigned GO terms. **(H)** Percentage of response-to-stimulus Gene Ontology terms across root cell populations in hybrid-enriched and parent-enriched gene sets. Lines connect values across populations, and colors distinguish hybrid- and parent-enriched genes. **(I)** Percentage of metabolic-process Gene Ontology terms across root cell populations in hybrid-enriched and parent-enriched gene sets. Lines and colors are defined as in **H**. Together, **G–I** show that parent-enriched genes retain stronger stress- and stimulus-responsive signatures, whereas hybrid-enriched genes show greater representation of metabolic programmes within shared cell-type contexts.



462

463 **Figure 3. Developmental and spatial associations of non-additive gene expression in the maize**
464 **root cortex.**

465 **(A)** UMAP visualization of cortex cells colored by three transcriptionally distinct cortex subclusters,
466 cortex early, cortex I mature and cortex II mature. **(B)** Group-specific projections of cortex cells onto the
467 shared UMAP embedding for the three parental genetic groups: non-stiff-stalk (NSS), tropical-line (TL)
468 and stiff-stalk (SS). In each panel, turquoise points represent parental inbred cells, red points represent
469 cells from the corresponding B73-derived hybrids and light-grey points show the remaining cortex cells.
470 The projections illustrate the distribution and relative representation of inbred and hybrid cells across
471 cortex developmental states. **(C)** Distribution of M_Score (left) and E_Score (right) across cortex cells.
472 Colors indicate the corresponding developmental gene-expression scores, calculated using the gene
473 sets described in the Methods. **(D)** Developmental trajectory of cortex cells reconstructed using
474 Monocle3 and displayed in UMAP space. Colors indicate inferred pseudotime. **(E)** Distribution of non-
475 additive gene (NAG) module scores across cortex cells in the same UMAP embedding. Colors represent

the aggregate expression score of genes classified as non-additively expressed, revealing variation in non-additive transcriptional activity among cortex states. **(F)** Hybrid-specific associations between NAG module scores and developmental scores. Each point represents one hybrid combination. The horizontal and vertical axes show Pearson's correlations of NAG scores with E_Score and M_Score, respectively. Colors distinguish hybrid combinations, and point sizes indicate the number of associated genes. Dashed lines mark zero correlation. **(G)** Association between the module score for NAGs exhibiting positive mid-parent heterosis (MPH) and E_Score. Each point represents one cell. The red line indicates a fitted linear regression; Pearson's correlation coefficient and its associated significance are shown (r = 0.774, P < 0.001). **(H)** Anatomical annotation of spatial transcriptomic spots in a representative hybrid maize root transverse section. Colors indicate cell-type assignments. MX, metaxylem. **(I)** Spatial distribution of the positive-MPH NAG module score across cortex layers. Cortex spots are colored by module score, and other cell types are shown in grey. **(J)** Distributions of positive-MPH NAG module scores across outer, middle and inner cortex layers in ten root sections. Violin plots show score distributions, with embedded box plots indicating medians and interquartile ranges. Pairwise comparisons used two-sided paired Wilcoxon tests on section-level medians, with Benjamini–Hochberg correction for multiple testing.

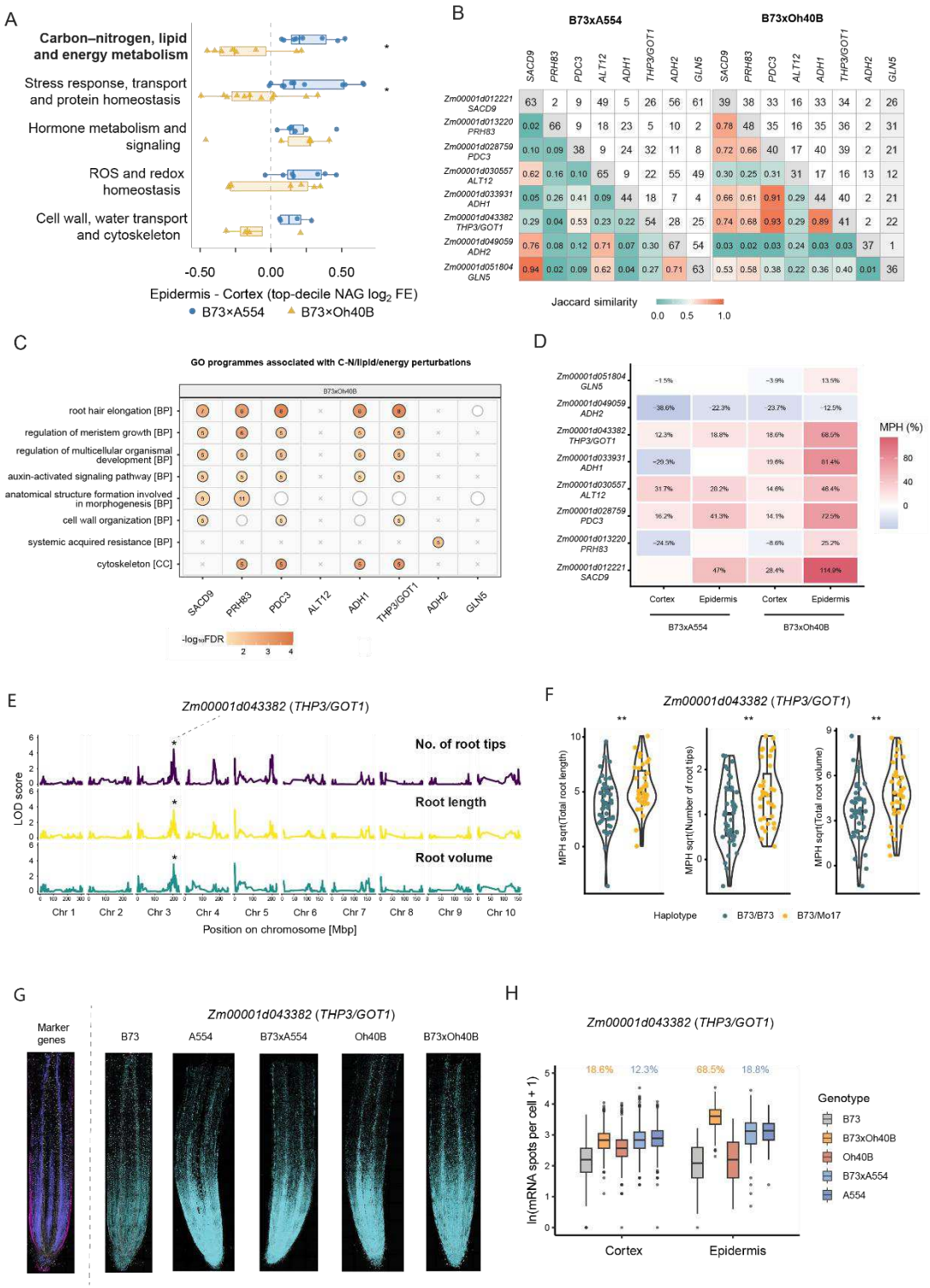


Figure 4. Metabolic and genetic evidence implicates a potential role of THP3/GOT1 in maize root heterosis.

(A) Tissue preference of predicted perturbation effects. The plot shows the epidermis-minus-cortex $\log_2$ fold enrichment of NAGs among the top 10% of genes ranked by predicted response to single-gene virtual knockout. Positive values indicate preferential enrichment in epidermis, whereas negative values indicate preferential enrichment in cortex. Colors denote five functional modules; points represent individual virtual KOs, with blue circles indicating B73 × A554 and orange triangles indicating B73 × Oh40B. Boxes show medians and interquartile ranges. **(B)** Shared and gene-specific effects of virtual knockout of eight carbon–nitrogen- and energy-associated candidates in B73 × A554 and B73 × Oh40B. Diagonal entries indicate the total number of significantly perturbed genes, upper-triangular entries show

shared number of perturbed genes between KO pairs, and lower-triangular entries show Jaccard similarity. **(C)** Gene Ontology enrichment among significantly perturbed genes in B73 × Oh40B. Rows show selected biological-process and cellular-component terms, and columns represent the eight virtual KOs. Circle size and internal labels indicate gene counts, whereas color indicates $-\log_{10}$-transformed false-discovery rate (FDR). **(D)** Cross- and tissue-specific expression heterosis of eight genes in the carbon–nitrogen, lipid and energy-metabolism module. Values indicate expression mid-parent heterosis (MPH) in cortex and epidermis of B73 × Oh40B and B73 × A554. Red and blue denote positive and negative MPH, respectively, while the displayed values show the exact MPH; blank tiles indicate non-significant MPH. The module includes the indicated genes. **(E)** Genome-wide QTL profiles for mid-parent heterosis (MPH) in root-tip number, total root length and total root volume. LOD scores are plotted against genetic position across the ten maize chromosomes. Asterisks indicate significant QTL peaks, including shared loci detected across the three root traits. **(F)** MPH values for total root length, root-tip number and total root volume in an IBM recombinant inbred line-derived population crossed with B73 and Mo17. Values are shown on a square-root-transformed scale and are stratified by genotype at the *THP3/GOT1* QTL position. The B73/B73 class represents the homozygous genotype, whereas the B73/Mo17 class represents the heterozygous genotype. Violin plots show the distribution of values, points represent individual genotypes and embedded box plots indicate the median and interquartile range. Colors denote genotype class. *P* values indicate comparisons between the homozygous and heterozygous genotype classes. The heterozygous B73/Mo17 class showed higher MPH for all three root traits than the homozygous B73/B73 class. **(G)** Spatial expression of *Zm00001d043382* (*THP3/GOT1*) in longitudinal roots of B73, B73 × Oh40B, Oh40B, B73 × A554 and A554. RNA-FISH gene-expression signals is shown across genotypes, revealing cross- and genotype-dependent regulations along the root developmental axis. **(H)** Tissue-resolved expression of *THP3/GOT1* in cortex and epidermis. Box plots show ln-transformed expression in the parental inbred lines and their corresponding hybrids. Values above hybrid groups indicate expression MPH for B73 × Oh40B (orange) and B73 × A554 (blue), respectively. Boxes indicate the interquartile range, central lines denote medians and whiskers extend to 1.5 times the interquartile range. *THP3* encodes glutamate–oxaloacetate transaminase 1 (GOT1), linking amino-acid metabolism with cellular carbon–nitrogen balance.

Methods

Plant materials and growth conditions

Maize (*Zea mays* L.) inbred lines and B73-based F1 hybrids were used to resolve cell-state-specific expression heterosis in primary roots. The single-cell experiment comprised three biological samples of B73, nine non-B73 paternal inbred lines and the nine corresponding F1 hybrids. The paternal lines represented three genetic groups: stiff-stalk (A680, B37 and NC328), non-stiff-stalk (A554, Oh40B and Oh43) and tropical (NC298, NC320 and NC352). The complete cross design, sample identifiers and parental relationships are provided in Supplementary Figure 6.

Unless otherwise specified, seeds were rinsed three to five times with ultrapure water, soaked in ultrapure water for 24 h and transferred to germination paper rolls positioned vertically in beakers containing double-distilled water. For the single-cell experiment and the anatomical phenotyping, 19 genotypes were grown for five days under control conditions at 26 °C/16 °C day/night with a 16-h light/8-h dark photoperiod. Roots used for LCM, spatial transcriptomics and targeted in situ sequencing were synchronized and collected at the same developmental stage i.e. 5-day old seedling. For all assays, the sampled root region and distance from the apex were standardized within an experiment.

Biological replicates were defined as independently grown plants or pooled roots from an independent growth unit, as indicated in the specific figure legends. The numbers of plants, roots, sections, cells or spatial observations, experimental blocks and sequencing libraries are reported accordingly. Plants were randomized within growth chambers or experimental blocks. Sample processing and image quantification were performed without knowledge of the expected heterosis class where practicable.

Root collection and preparation

For single-cell sequencing, approximately ten uniformly growing seedlings were selected for each preparation, and approximately 1-cm root tips were excised and processed immediately for protoplasting. For LCM and spatial transcriptomic assays, roots were collected, embedded in the appropriate tissue freezing medium (Polysciences, Warrington, USA, for LCM; Tissue-Tek® O.C.T.™ for spatial transcriptomic) and stored under -80°C freezer before downstream processing.

LCM sampling, library preparation and reference transcriptomes

Tissue isolation by laser-capture microdissection

LCM was used to obtain transcriptomes from anatomically defined root populations. The sampled populations were epidermis, outer cortex, inner cortex, endodermis, phloem-pole pericycle, xylem-pole pericycle, proto-phloem, proto-xylem, metaxylem, cells between metaxylem and pith. Root segments approximately 3–5 mm from the root tip were excised and fixed immediately using the RNA-preserving fixation procedure described by Yu *et al.* (2015).

After fixation, samples were vacuum-infiltrated with the Farmer's fixative (ethanol: acetic acid= 3:1), the fixative was replaced, and the vacuum-infiltration/swirl cycle was repeated twice. Seedling roots were vacuum-infiltrated at -500 mbar on ice for 15 min, followed by gentle swirling (20 rpm, neoLab, Heidelberg, Germany) at 4 °C for 1 h. The fixative was subsequently replaced with fresh sucrose

cryoprotectant (Sigma–Aldrich), applied to seedling roots in successive 10% and 15% steps. Vacuum infiltration continued until the tissue was uniformly infiltrated and protected against ice-crystal formation. Infiltrated root segments were embedded in 7 × 7 × 5-mm disposable molds (Polysciences, Warrington, USA) containing PolyFreeze-clear tissue-freezing medium (Polysciences, Warrington, USA), rapidly frozen and stored at −80 °C until sectioning. Cryosections were prepared using cryostat (CM1850, Leica Biosystems, Nussloch, Germany) and mounted on PEN membrane slide (Carl Zeiss, Göttingen, Germany). Tissue boundaries were identified by bright-field microscopy and anatomical landmarks. Regions of interest were dissected using a PALM MicroBeam laser-microdissection system (Carl Zeiss Microscopy; Jena, Germany) controlled by PALM RoboSoftware v4.8. Laser power, focus, cutting speed and collection mode were optimized on representative sections and recorded for each batch. Captured material was collected contact-free into adhesive cap containing lysis buffer.

Approximately 2,000 target cells or equivalent microdissected tissue area were collected for each population and biological replicate. Sections were processed rapidly under RNase-free conditions to minimize RNA degradation and cross-contamination between adjacent cell layers.

LCM RNA extraction, amplification and quality control

RNA was extracted, purified and quantified using Arcturus™ PicoPure™ RNA Isolation Kit (Thermo Fisher, Vilnius, Lithuania) under RNase-free conditions. RNA integrity was assessed using an Agilent 2100 Bioanalyzer with the RNA 6000 Pico Kit (Agilent Technologies, Boeblingen, Germany). Only samples with an RNA integrity number of at least 6.5 were advanced to amplification. For each sample, 20 ng of high-quality RNA was used for cDNA synthesis and amplification. Construction and sequencing of the paired-end mRNA libraries were performed on a DNASEQ platform at BGI.

LCM library construction, sequencing and read processing

cDNA libraries were prepared using DNBSEQ Low Input Smart-Seq Eukaryotic mRNA libraries. Libraries were randomized across sequencing lanes to minimize lane-associated confounding and sequenced using paired-end sequencing PE100bp on a DNBseq. Adapter sequences and low-quality bases were removed using SOAPnuke software (Chen et al., 2018) with parameters "-n 0.001 -l 20 -q 0.4 --adaMR 0.25 --polyX 50 --minReadLen 100" (v2.1.8). Clean reads were aligned to the maize B73 reference genome version 4 (http://ftp.ensemblgenomes.org/pub/plants/release-50/plants/fasta/zea_mays/dna/Zea_mays.B73_RefGen_v4.dna.toplevel.fa.gz) using HISAT2 (v2.2.1) (Kim et al., 2019) with default parameters. HISAT2 alignment output was piped directly to SAMtools (v1.5) (Li et al., 2009) and sorted by genomic coordinates to produce sorted BAM files for subsequent gene-expression analyses.

All downstream analyses were performed in R (v4.4.3). All BAM files were input to 'featureCounts' function (Liao et al., 2014) in R package Rsubread (v2.20.0) using maize B73 reference annotation v.4 (http://ftp.ensemblgenomes.org/pub/plants/release-50/gtf/zea_mays/Zea_mays.B73_RefGen_v4.50.chr.gtf.gz) to generate the gene expression table. Chimeric reads and reads mapped to more than one position in the genome were removed. Only genes

represented by a minimum of five mapped reads in ≥ 2 samples were declared as expressed and considered for downstream analyses.

LCM reference profiles and marker selection

Raw gene count matrices were analyzed using the R package edgeR (v4.4.2) (Robinson et al., 2010). Library size differences among samples were normalized using the trimmed mean of M-values (TMM) method. Normalized expression values were used for Multi-dimensional scaling plot using the plotMDS function in limma package (v3.62.2) (Ritchie *et al.*, 2015). Differential gene expression analysis was performed using the quasi-likelihood negative binomial generalized linear model framework implemented in edgeR (Robinson *et al.*, 2010). Dispersion parameters were estimated using the glmQLFit function, and differential expression was assessed using glmQLFTest function. *P*-values were adjusted for multiple testing using the Benjamini–Hochberg method. Unless otherwise stated, genes with an BH-adjusted *P*-value < 0.01 and an absolute $\log_2$ fold change ≥ 1 were considered differentially expressed (Benjamini and Hochberg, 1995). For each LCM population, differential expression was tested against the remaining LCM samples. Genes were ranked by log counts per million, and the 50 highest-ranking genes were retained as candidate reference markers for that population. These marker lists were used, together with anatomical position and published marker genes, to annotate spatial transcriptomic domains and single-cell clusters. The complete marker lists and test statistics are provided in Supplementary Table S1. All plots were produced using R package ggplot2 (v3.5.2) (Wickham, 2016)**.**

Single-cell RNA sequencing using the 10x Genomics platform

Root protoplasting and library preparation

Root tips were pretreated in 5 ml of 3% sorbitol solution containing 2.5 mM L-cysteine for 45 min at 26 °C to improve enzyme efficiency, followed by three washes with ultrapure water. Subsequently, root tips were finely chopped and transferred to 5 ml of protoplasting solution containing 20 mM KCl, 10 mM $CaCl_2$, 0.1% BSA (Sigma Aldrich), 20 mM MES, and 0.4 M Mannitol, adjusted to pH 5.7 with Tris HCl (pH=8.0). The digestion solution was supplemented with 1.2% cellulase 'Onozuka' RS, 1.2% cellulase 'Onozuka' R10, 0.4% macerozyme R-10 and 0.36% pectolyase Y-23 (all from Yakult Pharmaceutical) and protoplasting was performed at 26 °C with constant shaking at 60 rpm for 2 h. Protoplasts were then filtered through a 40-µm cell strainer (Corning®) followed by 3 min centrifuge at 100 g with acceleration and deceleration set to 3. After removing the supernatant, the protoplasts were resuspended in 0.5 M mannitol solution. Protoplast viability was assessed using trypan blue (Gibco™) and fluorescein diacetate (Invitrogen™) staining, and only cells with viability >70% were used for single cell library preparation. The protoplasts were quantified using a hemocytometer and adjusted to a density of approximately 800-1,200 cells/µl before being processed with the 10x Genomics Single Cell Protocol (CG000315, RevE).

Single cell RNA-seq libraries were prepared from freshly isolated protoplasts using Chromium Single Cell 3′ Reagent Kit V3.1 according to the manufacturer's instructions (10× Genomics) (Zheng et al., 2017). Specifically, 15,000 cells were loaded onto a single cell chip for library construction using the

10x Genomics Chromium single cell microfluidics device. Protoplasts were partitioned into single-cell GEMs (gel beads in emulsion), followed by cell lysis and barcoded reverse transcription of RNA within the droplets. Final library size and quality was checked on a DNA High Sensitivity Bioanalyzer chip (Agilent). The libraries were quantified using real-time quantitative PCR (TaqMan Probe), and sequencing was carried out using the DNBSEQ™ platform using PE100.

Read processing and cell-level quality control

Raw sequencing data were demultiplexed from Illumina BCL files using Cell Ranger mkfastq (version 6.1.2). FASTQ files were processed with Cell Ranger count/kallisto|bustools/scKB against the maize reference genome and gene annotation same as used for LCM RNA sequencing. Gene–cell count matrices were generated after barcode correction and UMI deduplication.

Single-cell transcriptomes from roots of all 21 samples were processed using Seurat (v5.3.0) (Hafemeister and Satija 2019). Cells with fewer than 1,000 unique molecular identifiers (UMIs), fewer than 500 detected genes, or greater than 5% mitochondrial transcript abundance were excluded from downstream analyses. Genes detected in fewer than three cells were also removed. Putative doublets were identified and removed using the scDblFinder package (v1.24.7) (Germain et al., 2021) with default parameters. After filtering, there are 91,157 cells and 27,768 genes kept for downstream analysis.

Following quality-control filtering, transcriptomes from three independent B73 biological replicates were compared to assess reproducibility. Pairwise correlation analyses revealed high correlations among replicates, indicating minimal technical or biological variation. Consequently, all three B73 replicates were retained for downstream analyses, and average gene expression across replicates was used as the parental B73 reference for comparative analyses and non-additive expression assessment.

Normalization, dimensionality reduction and dataset integration

To construct a unified single-cell atlas across genotypes, datasets from all 21 samples (three B73 replicates, 9 paternal inbred lines, and 9 hybrid samples) were normalized using the SCTransform workflow implemented in Seurat. Cross-sample integration was performed using the canonical correlation analysis (CCA)-based integration framework to minimize batch effects while preserving biologically meaningful variation. Principal component analysis was performed on the integrated dataset using RunPCA (npcs = 30), followed by graph-based clustering using a shared nearest-neighbor approach FindNeighbors (dims = 1:30) in Seurat. Graph-based clusters were identified at a resolution of 0.18 with first 30 PCs using the Louvain algorithm in FindClusters (algorithm = 1), resulting in 15 transcriptionally distinct cell populations.

Cell-type identities were assigned based on the expression of previously characterized maize root marker genes, spatial transcriptome references, and laser-capture microdissection (LCM)-derived marker datasets. Cluster annotations were further validated by consistency of marker gene expression patterns and agreement with established root cell-type transcriptional signatures. Uniform Manifold Approximation and Projection (UMAP) was used for visualization of cell populations and assessment of cluster separation.

The integrated single-cell atlas was subsequently used for cell-type-specific analyses of parental and hybrid transcriptomes, including the identification of non-additive gene expression patterns associated with heterosis (see below). To reduce false positive non-additive genes, non-additive genes were identified for genes that expressed at least 10% of cells in each cluster.

Cortex re-clustering and visualization of cell density

Following quality control, integration and cell-type annotation of the full scRNA-seq dataset, 33,763 cortex cells were extracted from the integrated Seurat object for joint analysis. These cells represented ten parental inbred lines and nine B73-derived F_1 hybrids. The RNA count matrix was re-normalized using Seurat::SCTransform with vst.flavor = "v2" and variable.features.n = 3,000 (Hafemeister and Satija, 2019). PCA was performed on the mean-centred Pearson residuals stored in the SCT assay, retaining 30 principal components. A nearest-neighbour graph and shared-nearest-neighbour graph were constructed using FindNeighbors with dimensions 1–20, \(k = 20\) and prune.SNN = 1/15 (Stuart *et al.*, 2019). Graph-based clustering was performed using the Louvain algorithm (algorithm 1; Blondel *et al.*, 2008) at a resolution of 0.04, yielding three cortex subclusters. A two-dimensional UMAP embedding was generated using the uwot implementation in RunUMAP with dimensions 1–15, cosine distance, 30 neighbours, min.dist = 0.3 and spread = 1 (McInnes *et al.*, 2018).

Cell-density plots were generated from the coordinates of the common cortex UMAP embedding. Local density was calculated independently for each group using ggpointdensity::geom_pointdensity with adjust = 1 (Kremer and Anders, 2025). All cells were retained, with each cell contributing one point. Density-coloured scatter plots were displayed against a black background using the magma palette; darker purple indicated lower local density, whereas orange yellow indicated higher local density. Colour scales were determined independently for each panel.

Pseudobulk analysis and sample-level comparisons

Pseudobulk expression profiles

To assess global transcriptional relationships among samples, single-cell expression profiles were aggregated to the sample level using Seurat's AggregateExpression function (Stuart *et al.*, 2019), grouping cells by sample identity. Highly variable genes were identified from the resulting pseudobulk profiles using FindVariableFeatures function, and principal component analysis was performed using the top 10 principal components with RunPCA function. The percentage of variance explained by each component was calculated from the corresponding eigenvalue-derived variance. PCA was used to assess relationships among B73, paternal lines and hybrids and to evaluate sample-level reproducibility.

Pseudobulk differential expression analysis

To characterize cell-type-specific co-expression modules and their association with hybrid–parent contrasts, pseudobulk expression profiles were generated for each cell type and analysed using hdWGCNA (Morabito *et al.*, 2023). An unsigned co-expression network was constructed, such that positively and negatively correlated genes could contribute to the same module based on the absolute strength of their pairwise correlations. Co-expression modules were identified according to the

hdWGCNA workflow, and module eigengenes were subsequently calculated to summarize the expression pattern of each module.

To identify modules showing differential activity between hybrids and parental inbred lines, module eigengenes were compared using the FindDMEs function in hdWGCNA. Hybrids and parents were treated as the two comparison groups, and differences in module eigengene values were assessed using the Wilcoxon rank-sum test. Modules with $\log_2$ fold change >1 and adjusted *P* value <0.05 were classified as hybrid-enriched modules, whereas modules with $\log_2$ fold change <(-1) and adjusted *P* value <0.05 were classified as parent-enriched modules. Multiple-testing correction was performed using Bonferroni correction.

All genes assigned to the respective hybrid- or parent-enriched modules were then pooled and subjected to Gene Ontology (GO) enrichment analysis to identify biological processes overrepresented among genes associated with each expression pattern (see below).

Developmental programmes and cortex trajectory

Meristematic and elongation-associated gene programmes

Developmental gene programmes were defined using the public maize root-slice RNA-seq dataset GSE172277, comprising 32 libraries from 16 consecutive longitudinal root positions, with two replicates per position. For this analysis, positions S1–S7 were assigned to the meristematic zone, S8 to the transition zone and S9–S16 to the elongation-associated zone.

Differential expression was analysed using DESeq2 (v1.48.2) (Love *et al.*, 2014), with developmental zone as the design factor. Genes with at least 10 counts in at least two libraries were retained, yielding 20,046 genes. Three comparisons were performed: transition versus elongation, meristematic versus elongation, and transition versus meristematic. Log_2 fold changes were shrunk using Deseq2::lfcShrink (type="apeglm"). Genes with a Benjamini–Hochberg-adjusted $P < 0.05$ and an absolute shrunken $\log_2$ fold change > 1 in at least one comparison were combined into a non-redundant set of 10,139 genes.

Variance-stabilizing transformation was applied to the filtered count matrix using vst (blind = TRUE) (Love *et al.*, 2014). Expression values for the selected genes were then standardized to gene-wise *z*-scores across all 32 libraries. Genes were grouped into four clusters using *k*-means clustering. Based on their expression profiles along the root developmental axis and published maize ground-tissue developmental signatures (Ortiz-Ramírez *et al.*, 2021), cluster 2 was defined as the elongation-associated programme (3,873 genes), whereas clusters 3 and 4 jointly defined the meristematic-associated programme (5,251 genes). Cluster 1 was excluded from both programmes.

The two programmes were scored separately across all 33,763 cortex cells using Seurat::AddModuleScore (Tirosh *et al.*, 2016; Stuart *et al.*, 2019), based on log1p-transformed, depth-corrected UMI counts from the SCT assay data slot generated using SCTransform (Hafemeister & Satija, 2019). Only programme genes present in this assay were included, yielding 5,100 meristematic-associated and 3,772 elongation-associated genes. The resulting scores, M_Score and E_Score, were visualized on the shared cortex UMAP using FeaturePlot.

Cortex trajectory and pseudotime

Cells were converted from the Seurat object into a Monocle3 cell_data_set using SeuratWrappers::as.cell_data_set and subsequently analysed with Monocle3 (Cao *et al.*, 2019).

The existing Seurat UMAP coordinates were transferred to Monocle3 object and used for trajectory inference. Monocle3 clustering and partition assignment were performed using cluster_cells (resolution = 0.001), followed by principal-graph learning within the resulting partitions using learn_graph. A root node was selected interactively in the lower UMAP region with high M_Score, thereby orienting the trajectory from a meristematic-associated transcriptional state. Pseudotime was assigned using order_cells and represents the distance along the learned graph from the selected root.

Comparison of cortex transcriptional distributions and relative parental similarity

Cortex cells were pooled into seven populations: the shared parent B73, and F1 and non-B73 parental (P1) populations for each of the non-stiff-stalk, stiff-stalk and tropical-line groups. Hybrids were assigned to the heterotic group of their non-B73 parent. Distributional comparisons used SCT-corrected UMI counts exported from the Seurat object.

A common expression representation was constructed for all populations. Counts were normalized to 10,000 per cell and log1p-transformed. From each population, 500 cells were randomly selected without replacement, yielding 3,500 training cells. Only these training cells were used to select 2,000 highly variable genes using Scanpy(v1.7.2) (Wolf *et al.*, 2018) and to fit gene-wise centering and scaling using StandardScaler (Pedregosa *et al.*, 2011). Scaled values above 10 were capped at 10. PCA was fitted on the training cells using the ARPACK solver (Lehoucq et al., 1998), retaining 10 principal components. The fitted scaling and PCA transformations were then applied to all cells.

A Gaussian radial basis function (RBF) kernel was used (Gretton et al., 2012). The bandwidth σ was estimated as the median pairwise Euclidean distance among the 3,500 training cells in PCA space (σ = 19.4431), and fixed across all comparisons. After excluding training cells, 2,000 cells were selected without replacement from each population to form fixed test sets (the same B73 test set shared across groups). Unbiased MMD² estimates were calculated between F1 and B73 and between F1 and P1 within each heterotic group. Relative parental similarity was evaluated by testing

$$\Delta = \mathrm{MMD}^2(F1, B73) - \mathrm{MMD}^2(F1, P1)$$

against zero using the authors' implementation of the relative MMD three-sample test, which accounts for covariance arising from the shared F1 sample. Two-sided *P* values were calculated as $\mathbf{2}\boldsymbol{\Phi}(-|\boldsymbol{z}|)$, where z is the standardized relative-MMD statistic, and Bonferroni correction was applied across the three heterotic groups (Bounliphone et al., 2016).

For visualization, 100 cells were sampled with replacement from each fixed F1, B73 and P1 test population over 100 iterations. Within each iteration, both parental comparisons used the same F1 subsample and the fixed kernel. The resulting MMD² estimates were summarized as boxplots; significance annotations were obtained from the relative MMD test on the full test sets.

Non-additive expression and heterosis calculations

Definition of non-additive genes

Non-additive expression was evaluated within the same cell type or developmental state by comparing hybrid expression with the mean expression of the two corresponding parents (Swanson-Wagner *et al.*, 2006). For a gene g, the parental midpoint was calculated as:

$$MP_g = \frac{P_{1g} + P_{2g}}{2},$$

where P_{1g} and P_{2g} denote the expression levels of gene g in the two parental materials. Expression mid-parent heterosis was calculated as:

$$MPH_g(\%) = \frac{F_{1g} - MP_g}{MP_g} \times 100 = \left(\frac{F_{1g}}{MP_g} - 1\right) \times 100,$$

The expression scale used for these calculations was normalized expression. Genes were tested only when they were expressed in at least 10% of cells in the relevant population. Hybrid–parent contrasts were estimated with two-sided Student's t-tests, and adjusted *P* value using the Bonferroni procedure. A gene was classified as non-additive when hybrid expression significantly deviated from the parental midpoint at corrected *P* <0.05 implemented with the *p*.adjust() function in R.

The analysis key was hybrid × cell type × gene. Positive-MPH NAGs were defined as NAGs with $MPH_g > 0$, and negative-MPH NAGs as NAGs with $MPH_g < 0$. Genes with inconsistent directions across the selected annotations were excluded from directional module analyses.

NAG recurrence and regulatory modes

For each cell type and hybrid, the NAG proportion was calculated as the number of NAGs divided by the number of expressed genes tested in that population. A gene was considered cell-type-specific when it was classified as non-additive in one cell type but not in the other tested cell types. Recurrence was defined as the number of hybrid combinations in which a gene was classified as non-additive.

Regulatory modes were assigned by comparing F1 expression with the two parental values. Genes were classified as over-dominant, under-dominant, maternal-dominant and paternal-dominant based on the previous published method (Swanson-Wagner *et al.*, 2006).

Spatial transcriptomic sample preparation and sequencing

Tissue preparation and Stereo-seq library construction

Spatial transcriptomics was performed on primary-root segments from B73, A554 and B73 × A554. Approximately 5-cm root segments extending from the root tip were excised, and 12 segments per genotype were embedded in Tissue-Tek O.C.T. compound in 1 cm × 1 cm Cryomold containers matched to the dimensions of the Stereo-seq chip. Samples were frozen immediately in liquid nitrogen and stored at −80 °C. Transverse cryosections 10 μm thick were prepared using a Leica CM1950 cryostat. RNA integrity was assessed using an Agilent Bioanalyzer, and samples with an RNA integrity number greater than 9.5 were used for spatial library preparation.

Before the formal experiment, serial sections were used to assess RNA integrity, tissue morphology and the optimal permeabilization time. Sections mounted on ordinary slides were methanol-fixed and stained with toluidine blue for anatomical inspection. For permeabilization optimization, sections were placed on test chips, exposed to a time gradient of permeabilization reagent and evaluated by fluorescent cDNA signal. For the formal experiment, sections were placed on spatially barcoded Stereo-seq chips, incubated at 37 °C for 3 min to promote adhesion, fixed in methanol at −20 °C for 30 min and stained with a nucleic-acid dye (Thermo Fisher Scientific, Q10212) for imaging. Stereo-seq libraries were prepared with the STOmics Gene Expression Kit S1 (BGI, 1000028493) according to the manufacturer's protocol and the published Stereo-seq workflow (Wang *et al.*, 2025).

The spatial workflow comprised tissue permeabilization, in situ reverse transcription, removal of residual tissue, release and recovery of chip-bound cDNA, cDNA purification and amplification, fragmentation, PCR-based library construction, dual-size selection and library quality control. Amplified libraries were circularized, amplified by rolling-circle amplification to generate DNA nanoballs and sequenced using patterned nanoarrays and cPAS. Paired-end sequencing was performed on a DNBSEQ-T1, MGISEQ-2000 platform, using 50-bp read 1 and 100-bp read 2, as specified by the sequencing batch.

Spatial read processing and quality control

Raw sequencing data were processed using the SAW pipeline (Stereo-seq Analysis Workflow, v7.1.2). Briefly, sequencing reads were demultiplexed, spatial barcodes and unique molecular identifiers (UMIs) were extracted, and low-quality reads were removed. Clean reads were aligned to the maize B73 reference genome (B73 RefGen_v4; Ensembl Plants release 50) using the corresponding genome assembly and gene annotation files (see above). Gene expression matrices were generated by assigning uniquely mapped and annotated transcripts to their spatial coordinates, followed by UMI deduplication. Spatial bins were subsequently aggregated at the specified bin size (Bin50 = 25 μm × 25 μm) to generate spatially resolved gene-expression profiles for downstream analyses.

To establish a reference atlas for single cell-type annotation, spatial transcriptome data from 12 root sections of maize inbred line B73 were analyzed together in Seurat (v5.3.0). In total, 12,256 spatial bins containing expression profiles for 28,843 genes were obtained. Quality control filtering removed bins expressing fewer than 200 genes and genes detected in fewer than three bins. Two sections (G1 and G9) were excluded from further analyses because of substantially lower cell recovery compared with the remaining sections. The remaining ten sections were integrated using Seurat and clustered at a resolution of 1.7. One cluster corresponding to spatial bins located outside the tissue boundary, as determined by cDNA image inspection, was removed. The final B73 reference atlas consisted of 9,832 spatial bins and 16,366 expressed genes.

Spatial transcriptomic data analysis

To investigate non-additive gene expression, spatial transcriptome datasets from the three genotypes were jointly analyzed. Spatial bins and genes were filtered using the same quality control criteria described above. Following filtering, a total of 16,383 genes across 34,463 spatial bins were retained for downstream analyses. Gene expression values were normalized using the SCTransform workflow

implemented in Seurat. Batch correction and dataset integration were performed using canonical correlation analysis (CCA)-based integration. The integrated dataset was clustered using a resolution of 1.2, resulting in 20 transcriptionally distinct clusters which were grouped to 11 cell types based on expression of previously characterized maize root marker genes, laser-capture microdissection (LCM)-derived markers and spatial positions. Dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP) implemented in RunUMAP with first 30 PCs, and cluster distributions were visualized using both DimPlot and SpatialDimPlot functions in Seurat.

Non-additive gene expression was subsequently assessed within individual cell types by comparing hybrid expression levels with the corresponding parental expression profiles (see below). Spatial–single-cell NAG overlap was calculated as the proportion of shared genes compared to NAGs in spatial transcriptome. Because spatial bins can contain transcripts from neighbouring cells, overlap was interpreted as support for cell-type-associated programmes rather than as a one-to-one measurement of single-cell expression.

Gene Ontology enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using the enricher function implemented in the R package clusterProfiler (v4.14.6) (Wu et al., 2021). Maize GO annotations were obtained from the B73 AGPv4 genome annotation in GAF format and used to construct a custom TERM2GENE annotation table linking GO identifiers to maize gene models. GO term names, ontology categories and definitions were retrieved from the GO.db package (v3.20.0).

For each cluster, genes classified as non-additive expressed or differentially expressed were used as the query gene set. Enrichment was tested against all genes in the annotation file. *P* values were adjusted for multiple testing using the Benjamini–Hochberg method, and GO terms with an adjusted *P* value < 0.05 were considered significantly enriched.

Multiplex *in situ* spatial validation

Candidate genes were selected by integrating cell-type-enriched expression, non-additive gene (NAG) status, functional annotation, spatial transcriptomic localization and their relevance to the identified heterosis-associated programmes. Multiplex *in situ* spatial validation was performed for 8 candidate genes linked to carbon and nitrogen metabolism using a modified MiP-seq protocol described previously by Wu et al., 2024.

Tissue preparation

Maize plants of B73, A554, Oh40B, B73 × A554 and B73 × Oh40B were grown under the conditions described above. Root samples were collected from 5-day-old maize seedlings, washed carefully with deionized water and immediately fixed by perfusion with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS). Samples were subsequently post-fixed in 4% PFA for 12 h at 4 °C.

For cryoprotection, fixed roots were incubated in 30% sucrose prepared in PBS until the tissue was fully equilibrated. Samples were embedded in optimal cutting temperature (OCT) compound and sectioned serially at 10 μm using a Leica cryostat (LeicaCM1950). Sections containing the complete root tip or the

specified longitudinal and transverse anatomical regions were collected and stored at −80 °C until *in situ* processing.

Before probe hybridization, sections were enclosed in Secure-Seal hybridization chambers and rinsed twice with ice-cold diethyl pyrocarbonate (DEPC)-treated PBS. Sections were re-fixed in 4% PFA for 10 min and washed three times with DEPC-treated OBSTR buffer consisting of PBS supplemented with 0.1% Tween-20 and 0.1 U μl^{-1} RNase inhibitor. Sections were then treated with pre-cooled methanol at −80 °C for 15 min. Permeabilization was performed with 2 mg ml^{-1} pepsin (Sigma-Aldrich, P0525000) in 0.1 M HCl at 37 °C for 90 s. Sections were rinsed with DEPC-treated PBSTR buffer before probe hybridization.

Probe design

For each candidate gene, two target sites were selected preferentially within 5′ exonic regions. The probe set consisted of gene-specific padlock probes, rolling-circle amplification (RCA) initiator probes, RCA detection probes and sequencing probes. Padlock probes contained complementary target-binding arms of 12–16 bp, an internal anchor-primer sequence and a gene-specific barcode sequence. The RCA initiator probe contained a 12–16-bp target-complementary sequence and a region complementary to the padlock probe. Probe specificity and amplification efficiency were evaluated using the corresponding detection probe, including the Cy3-labelled cy3-p4 probe. In situ validation used three anchor primers and four query probes (seg-1, seg-2, seg-3 and seg-4), each labelled with distinct Alexa Fluor dyes.

Padlock-probe hybridization and rolling-circle amplification

Padlock probes were phosphorylated at their 5′ termini using T4 polynucleotide kinase (Vazyme, N102-472-01). Phosphorylated padlock probes and the corresponding RCA initiator probes were each used at a final concentration of 30 nM per probe in hybridization buffer containing 2× SSC, 10% formamide and 20 mM RVC. The hybridization mixture was applied to the pretreated sections and incubated overnight at 37 °C.

After hybridization, sections were washed twice with DEPC-treated PBSRT buffer and once with 4× SSC. Circularization of hybridized padlock probes was performed using a ligation reaction mixture for 2 h at 37 °C. Sections were then washed twice with DEPC-treated PBSTR buffer.

Rolling-circle amplification was performed for 2 h at 30 °C in a reaction mixture containing 1 U μl^{-1} Phi29 DNA polymerase (Vazyme, N106-01), 1× RCA buffer, 0.25 μM dNTPs, 0.2 μg μl^{-1} bovine serum albumin and 5% glycerol. Following amplification, sections were washed before the sequencing reaction.

In situ sequencing

The sequencing reaction mixture contained 1× BSA, 0.2 U μl^{-1} T4 DNA ligase (Thermo Fisher Scientific, EL0012), 1× T4 DNA ligase buffer, 1 μM anchor primer and 1 μM query primer. Sections were incubated with the sequencing mixture for 2 h at 25 °C. After each sequencing reaction, sections were washed three times with 10% formamide in 2× SSC and counterstained with DAPI.

Multiple sequencing rounds were performed to resolve the encoded candidate-gene identities. Following each imaging round, fluorescent sequencing signals were removed by two 10-min washes with stripping buffer containing 60% formamide in 2× SSC at room temperature. The next sequencing mixture was then applied to the sections.

Image acquisition and processing

Images were acquired using a Leica THUNDER Imaging System (Imager 3D Assay) equipped with a 20× objective (NA = 0.80) and the instrument's fluorescence detection system. Imaging settings, exposure times, laser or illumination intensities and detector parameters were kept constant among comparable genotypes and sections.

Image reconstruction, registration of sequential sequencing rounds and signal decoding were performed using ImageJ. DAPI staining was used to identify nuclei and to register images across sequencing rounds. Gene-specific signals were assigned according to the corresponding barcode and probe identity. Fluorescence images acquired were registered in ImageJ using BigWarp (Bogovic et al., 2016), which applies thin-plate spline–based elastic transformations within the BigDataViewer (Pietzsch et al., 2015) framework. Following registration, RNA signals were detected and decoded according to their combinatorial fluorescence codes. RNA spots were identified using RS-FISH and U-FISH, with RS-FISH used for spot localization and intensity extraction and U-FISH used for deep-learning-based spot enhancement and detection through a U-Net architecture (Bahry *et al.*, 2022; Xu *et al.*, 2025). Cell boundaries were segmented using Cellpose v3.0 (Stringer *et al.*, 2021). Decoded transcript coordinates were subsequently assigned to individual segmented cells using ImageFlow, which integrates transcript and cell-boundary coordinates through KD-tree–based spatial indexing, radius-adaptive nearest-neighbor searches, and probabilistic assignment scoring. The resulting transcript-to-cell assignments were used to construct single-cell spatial expression profiles for downstream analyses.

Spatial quantification

Spatial images and cell segmentation outputs generated by the sequencing platform were subsequently analyzed in Python (v3.9.21) for cell-type classification and quantitative assessment of transcript abundance.

To identify epidermal and cortex cell populations within root sections, a combined marker gene- and geometry-based classification strategy was applied. First, cell types were preliminarily assigned based on the expression intensity of established tissue-specific marker genes in our single cell transcriptome data, including the cortex marker gene *Zm00001d019565* and the epidermis marker gene *Zm00001d037547*. Thresholds for marker expression were empirically determined based on signal distributions and spatial localization patterns across sections.

Second, spatial geometric information was incorporated to further refine tissue classification. Cell centroids were extracted from the segmentation masks, and the Euclidean distance between each cell centroid and the geometric center of the root section was calculated. To account for variation in root size and morphology among samples, distances were normalized within each section. Cortex cells were subsequently defined as cells located between the 40th and 90th percentile of the radial distance

distribution. Epidermal cells were identified as cells directly contacting the outer boundary of the root section.

Cell segmentation masks and boundary extraction were processed using the scipy ndimage module (v1.10.1), which was used for cell edge detection and spatial boundary identification. Cells located within the root cap region were excluded from all downstream analyses to avoid interference from highly heterogeneous terminal tissues. Spatial classification maps were visualized using matplotlib module (v3.7.1) in Python.

To quantify heterosis of gene expression within specific cell types, cell-level expression matrices generated from the Python-based spatial transcriptomic analysis were imported into R for downstream statistical analyses. Mid-parent heterosis (MPH) of gene expression was calculated by comparing the expression level of each hybrid within a given cell type to the mean expression value of its two parental lines (see above). Statistical significance of expression differences between hybrids and mid-parent values was assessed using two-sided *t*-tests implemented with the *t*.test() function in R. *P* values were adjusted by the Benjamini–Hochberg procedure where applicable.

Root anatomical traits and heterosis measurements

Primary roots from 19 genotypes were embedded in 8% (w/v) low-melting-point agarose prepared in water. Agarose blocks were trimmed and mounted on a Leica VT1200 S vibratome. Transverse sections 80 μm thick were collected, and sections obtained approximately 4 cm from the root tip were imaged using a Zeiss PALM MicroBeam microscope.

To quantify cell size across different root tissues, a three-step image analysis pipeline was established. First, raw cross-sectional images were processed in ImageJ (v1.54p) (Stringer and Pachitariu 2025) to enhance local contrast and improve cell boundary visibility using the “Enhance Local Contrast” function. This preprocessing step facilitated subsequent cell segmentation, particularly in tissues with weak cell wall contrast. Second, the enhanced images were subjected to automated cell segmentation using Cellpose (Schneider *et al.*, 2012). Cellpose identified individual cells based on contrast differences along cell edges and generated segmented masks for downstream quantitative analysis. Segmentation outputs were manually inspected to ensure consistent cell boundary detection across samples. Third, segmented images were imported into R for tissue classification and cell size quantification using package tiff (v0.1.12) and EBImage (v4.52.0). Cell size was calculated based on the segmented cell area using computeFeatures.shape function in EBImage package. To distinguish different cell layers within the root cross-sections, the Euclidean distance between each cell centroid (calculated with computeFeatures.moment function in EBImage package) and the geometric center of the root was computed. Distance thresholds were empirically adjusted for each sample according to visualization of the radial tissue organization in R, enabling classification of major tissue types along the radial axis.

Due to the relatively low contrast and small size of certain cell layers, particularly pericycle cells, accurate discrimination from adjacent tissues was not consistently achievable across all genotypes. Therefore, quantitative analyses were restricted to tissues that could be robustly identified, including pith, metaxylem, endodermis, inner cortex, middle cortex, outer cortex, and epidermis.

Mean cell size was subsequently calculated for each cell type in each sample. For cell size of each anatomical trait, mid-parent heterosis was calculated as:

$$MPH \quad (\%) = \frac{F_1 - (P_1 + P_2)/2}{(P_1 + P_2)/2} \times 100$$

Where F_1 denotes the cell size in hybrids, P_1 and P_2 denote the cell size in the two parental materials. Differences from the mid-parent expectation were tested using two-sided Student's t-test at biological-replicate level. *P* values were adjusted for multiple testing using the Benjamini–Hochberg procedure, and adjusted *P* <0.05 were considered statistically significant.

Associations between number of NAGs and anatomical heterosis were evaluated across the nine hybrids using Spearman correlation.

Virtual perturbation and functional analysis

Single-gene virtual knockout and NAG enrichment

Single-gene virtual knockouts were performed for 36 candidates from five functional modules in B73 × A554 and B73 × Oh40B using scTenifoldKnk(v2.0.0) (Osorio *et al.*, 2022). Hybrid-specific gene-by-cell matrices were extracted from the RNA count layer of the annotated Seurat object, retaining variable features identified by SCTransform and supplementing them with knockout targets. Networks were constructed using all cells from each hybrid, rather than separately for individual tissues. Ten networks were reconstructed from 500-cell subsamples using three principal components, an edge-weight quantile threshold of 0.9 and scaled, asymmetric weights. Rank-three tensor decomposition was performed with a maximum of 1,000 iterations and a convergence tolerance of 10^{-5}. Knockouts were simulated by zeroing the target gene's outgoing weights. Wild-type and knockout networks were compared by manifold alignment, and genes were ranked by the Z statistic (network displacement). Significance was defined as BH-adjusted *P* < 0.05.

For NAG enrichment, each knockout result was matched to the corresponding hybrid- and cell-state-specific additive/non-additive gene annotations. The knockout target and genes with non-finite Z statistics were excluded. Genes with average-rank percentiles ≥0.90 formed the high-perturbation set, without an additional significance filter. Enrichment was calculated as

$$L_{hkc} = log_2\left[\frac{(a + 0.5)/(N_{\text{NAG}} + 1)}{(b + 0.5)/(N_{\text{Additive}} + 1)}\right]$$

where h, k and c denote hybrid, knockout and cell state; a and b are the numbers of NAGs and additive genes in the high-perturbation set; and N_{NAG} and $N_{Additive}$ are their respective totals in the matched background. Positive values indicate preferential representation of NAGs among strongly perturbed genes.

Cortex enrichment was summarized by the median across the four cortex states, and tissue preference was defined as

$$D_{hk} = L_{hk,\text{Epidermis}} - \text{median}_{c \in C} L_{hkc}$$

where $\mathcal{C}$ comprises Cortex outer, Cortex II mature, Cortex I mature and Cortex early. Positive and negative values indicate epidermal and cortex preference, respectively.

Shared perturbation effects and Jaccard similarity

For each hybrid and knockout, $S_{\{h,k\}}$ comprised all significantly perturbed genes, excluding that knockout's target. Within each hybrid, overlap between two knockouts was quantified as

$$J_{h,ij} = \frac{\left|S_{h,i} \cap S_{h,j}\right|}{\left|S_{h,i} \cup S_{h,j}\right|}$$

Each knockout retained its original tested-gene background. Matrix diagonals report $\left|S_{\{h,k\}}\right|$, upper triangles report intersection counts and lower triangles report Jaccard similarity.

GSEA and GO enrichment

Pre-ranked GSEA was performed using fgsea (v1.38.0, Korotkevich *et al.*, 2016) for the eight carbon–nitrogen/lipid/energy candidates. For each knockout–context combination, the knockout target was excluded, and genes were matched to the corresponding cell-state annotation table and ranked by decreasing Z statistic. Enrichment was assessed using fgseaMultilevel (minSize = 10, eps = 1e-10). All NAGs, positive-MPH NAGs and negative-MPH NAGs were tested separately in B73 × A554 epidermis and the four B73 × Oh40B cortex states. *P* values were Benjamini–Hochberg-adjusted within each NAG category across the 40 knockout–context combinations (8 candidates × 5 contexts), with significance defined as FDR < 0.05.

GO over-representation analysis was performed using clusterProfiler (version 4.14.6) for the eight candidates in both hybrids, with a fixed B73 RefGen_v4 gene–GO mapping assembled from MaizeGAMER and MaizeGDB annotations. For each hybrid–knockout combination, the query comprised significantly perturbed genes and the knockout target; the background comprised all genes with finite adjusted *P* values from perturbation analysis and the target, restricted to genes represented in the GO mapping. Biological Process, Cellular Component and Molecular Function terms were tested using enricher (minGSSize = 5, pAdjustMethod = "BH", ont = c("BP","CC","MF"), pvalueCutoff = 0.05, qvalueCutoff = 0.05).

QTL mapping and local genotype association of root-trait heterosis

Root-tip number, total root length and total root volume were measured in the B73 × IBM-RIL population. To meet the assumptions of linear regression modelling, the trait values were transformed using the square root before estimating the mean for each genotype and trait (Pitz et al., 2025). All subsequent analyses were performed on the transformed scale. Trait values were converted to mid-parent heterosis (MPH) using the parental midpoint calculated for each corresponding cross. Genome-wide QTL scans were performed with R/qtl2 (version 0.22), incorporating conditional genotype probabilities calculated from a set of high-confidence SNP loci (Pitz et al., 2025). Genome-wide significance thresholds (α = 0.05) were determined by 1000 permutations, and significant QTL peaks were identified using the

corresponding LOD threshold. This analysis identified shared QTL signals on chromosomes 3 and 5 for the three root traits, including a chromosome-3 interval containing the prioritized *THP3/GOT1* locus.

To test the association between the chromosome-3 locus and root-trait heterosis, individuals were classified according to their genotype at the *THP3/GOT1* position as B73/B73 homozygotes or B73/Mo17 heterozygotes. Differences between the two classes were tested using Student's *t*-test, with significance indicated by asterisks codes (* = 0.05, **=0.001, *** = 0.0001).

Code availability

All custom code used for sequencing-data preprocessing, single-cell and spatial transcriptomic integration, cell-type annotation, non-additive expression analysis, developmental trajectory inference, virtual perturbation, QTL analysis and figure generation is available with link: https://github.com/Danning24/single-cell-and-spatial-transcriptome-analysis-of-heterosis. The repository contains analysis scripts, notebooks, parameter settings, software-version information, random seeds and instructions for reproducing the reported analyses. Source Data files provide the input tables and processed outputs required to reproduce the main and supplementary figures.

Data availability

The raw and processed single-cell RNA-sequencing, spatial transcriptomic and laser-capture microdissection RNA-sequencing datasets generated in this study have been deposited in the NCBI Gene Expression Omnibus under accession numbers GSE354594 and GSE354821. The repository includes the raw sequencing files, processed gene–cell or spot-by-gene matrices, sample metadata, cell-type annotations and spatial coordinates.

Source data for all figures, root anatomical measurements, mid-parent heterosis calculations, QTL genotypes and phenotypes, candidate-gene associations, virtual perturbation results and probe sequences are provided in the Supplementary Information and Source Data files. The complete metadata for the single-cell and spatial transcriptomic datasets are provided in Supplementary Tables 2 to 5, respectively. Probe sequences used for spatial in situ sequencing are provided in Supplementary Table 12, and QTL-linked candidate-gene information is provided in Supplementary Table 13.

An interactive browser for exploring the single-cell, spatial transcriptomic non-additive gene datasets is available at http://93.93.112.125/. Additional information is available from the corresponding author upon reasonable request.

References

Appels, M. A. & Haaker, H. Glutamate oxaloacetate transaminase in pea root nodules: participation in a malate/aspartate shuttle between plant and bacteroid. *Plant Physiol.* **95**, 740–747 (1991).

Bahry, E. *et al.* RS-FISH: precise, interactive, fast and scalable FISH spot detection. *Nat. Methods* **19**, 1563–1567 (2022).

Baldauf, J. A. *et al.* Single-parent expression complementation contributes to phenotypic heterosis in maize hybrids. *Plant Physiol.* **189**, 1625–1638 (2022).

Baldauf, J. A. *et al.* Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. *Curr. Biol.* **28**, 431–437 (2018).

Baldauf, J. A., Marcon, C., Paschold, A. & Hochholdinger, F. Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive, and allelic expression patterns in maize hybrids. *Plant Physiol.* **171**, 1144–1155 (2016).

Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Birchler, J. A., Auger, D. L. & Riddle, N. C. In search of the molecular basis of heterosis. *Plant Cell* **15**, 2236–2240 (2003).

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D. & Veitia, R. A. Heterosis. *Plant Cell* **22**, 2105–2112 (2010).

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of calcium images by learned contrast synthesis. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 1123–1126 (IEEE, 2016).

Bounliphone, W., Belilovsky, E., Blaschko, M. B., Antonoglou, I. & Gretton, A. (2016). A test of relative similarity for model selection in generative models. *International Conference on Learning Representations (ICLR 2016)*.

Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**, gix120 (2018).

Denyer, T. & Timmermans, M. C. P. Crafting a blueprint for single-cell RNA sequencing. *Trends Plant Sci.* **27**, 92–103 (2022).

Dolan, L. Root hair development in grasses and cereals (Poaceae). *Curr. Opin. Genet. Dev.* **45**, 76–81 (2017).

Duvick, D. N. Biotechnology in the 1930s: the development of hybrid maize. *Nat. Rev. Genet.* **2**, 69–74 (2001).

East, E.M. (1908) Inbreeding in corn. In Reports of the Connecticut Agricultural Experiment Station for Years 1907–1908. (East, E.M., ed), pp. 419–428, Connecticut Agricultural Experiment Station.

Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using *scDblFinder*. *F1000Research* **10**, 979 (2021).

Gouran, M. & Brady, S. M. The transcriptional integration of environmental cues with root cell type development. *Plant Physiol.* **196**, 2150–2161 (2024).

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. & Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research* **13**, 723–773.

Guillotin, B. *et al.* A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* **617**, 785–791 (2023).

Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

Hochholdinger, F. & Yu, P. Molecular concepts to explain heterosis in crops. *Trends Plant Sci.* **30**, 95–104 (2025).

Huang, Y. *et al.* Teosinte alleles enhance nitrogen assimilation and seed protein in maize. *Nature* (2026a).

Huang, Y. *et al.* Chromatin accessibility landscape and its association with heterosis in maize hybrids. *Nat. Commun.* **17**, 5659 (2026b).

Hoecker, N., Keller, B., Piepho, H.-P. & Hochholdinger, F. Manifestation of heterosis during early maize (*Zea mays* L.) root development. *Theor. Appl. Genet.* **112**, 421–429 (2006).

Hoecker, N., Keller, B., Muthreich, N., Chollet, D., Descombes, P., Piepho, H.-P. & Hochholdinger, F. Comparison of maize (*Zea mays* L.) F1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* **179**, 1275–1283 (2008).

Janko, K. *et al.* Unifying framework explaining how parental regulatory divergence can drive gene expression in hybrids and allopolyploids. *Nat. Commun.* **15**, 8714 (2024).

Jiang, L. *et al.* Cell-type-specific regulatory variation shapes maize heterosis. *bioRxiv* https://doi.org/10.64898/2026.08.20.744945 (2026).

Kajala, K. *et al.* Innovation, conservation, and repurposing of gene function in root cell type development. *Cell* **184**, 3333–3348 (2021).

Kim, D. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

Korotkevich, G., Sukhov, V., Golikov, N. & Sergushichev, A. *fgsea: Fast Gene Set Enrichment Analysis*, R package v1.38.0. Bioconductor. https://doi.org/10.18129/B9.bioc.fgsea

Kremer, L. P. M. & Anders, S. ggpointdensity: a cross between a 2D density plot and a scatter plot. R package v0.2.1 (2025). https://doi.org/10.32614/CRAN.package.ggpointdensity

Lehoucq, R. B., Sorensen, D. C. & Yang, C. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. (SIAM, 1998).

Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

Lippman, Z. B. & Zamir, D. Heterosis: revisiting the magic. *Trends Genet.* **23**, 60–66 (2007).

Liu, X., Hu, B. & Chu, C. Nitrogen assimilation in plants: current status and future prospects. *J. Genet. Genomics* **49**, 394–404 (2022).

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Lynch, J. P. Harnessing root architecture to address global challenges. *Plant J.* **109**, 415–431 (2022).

McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 (2018).

Minne, M., Ke, Y., Saura-Sanchez, M. & De Rybel, B. Advancing root developmental research through single-cell technologies. *Curr. Opin. Plant Biol.* **65**, 102113 (2022).

Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep. Methods* **3**, 100498 (2023).

Nolan, T. M. *et al.* Brassinosteroid gene regulatory networks at cellular resolution in the Arabidopsis root. *Science* **379**, eadf4721 (2023).

Omary, M. *et al.* A conserved superlocus regulates above- and belowground root initiation. *Science* **375**, eabf4368 (2022).

Orman-Ligeza, B. *et al.* Post-embryonic root organogenesis in cereals: branching out from model plants. *Trends Plant Sci.* **18**, 459–467 (2013).

Ortiz-Ramírez, C. *et al.* Ground tissue circuitry regulates organ complexity in maize and Setaria. *Science* **374**, 1247–1252 (2021).

Osorio, D. *et al.* scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns* **3**, 100434 (2022).

Paschold, A. *et al.* Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. *Genome Res.* **22**, 2445–2454 (2012).

Paschold, A. *et al.* Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell* **26**, 3939–3948 (2014).

Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. BigDataViewer: visualization and processing for large image data sets. *Nat. Methods* **12**, 481–483 (2015).

Pitz, M., Baldauf, J. A., Piepho, H. P. & Hochholdinger, F. Non-additive gene expression contributing to heterosis in partially heterozygous maize hybrids is predominantly regulated from heterozygous regions. *New Phytol.* **247**, 669–683 (2025).

Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Rogers, E. D. & Benfey, P. N. Regulation of plant root system architecture: implications for crop advancement. *Curr. Opin. Biotechnol.* **32**, 93–98 (2015).

Roszak, P. *et al.* Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization. *Science* **374**, eaba5531 (2021).

Schnable, P. S. & Springer, N. M. Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* **64**, 71–88 (2013).

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

Shull, G. H. The composition of a field of maize. *J. Hered.* **4**, 296–301 (1908).

Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *Nat. Methods* **22**, 592–599 (2025).

Stringer, C. *et al.* Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. III, Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21.

Swanson-Wagner, R. A. *et al.* All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl Acad. Sci. USA* **103**, 6805–6810 (2006).

Tian, H., De Smet, I. & Ding, Z. Shaping a root system: regulating lateral versus primary root growth. *Trends Plant Sci.* **19**, 426–431 (2014).

Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

Wang, X. et al. A single-cell multi-omics atlas of rice. *Nature* **644**, 722–730 (2025).

Wang, Z., Ni, Z., Wu, H., Nie, X. & Sun, Q. Heterosis in root development and differential gene expression between hybrids and their parental inbreds in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **113**, 1283–1294 (2006).

Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd edn (Springer, 2016).

Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

Wu, T. *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

Wu, X. et al. Spatial multi-omics at subcellular resolution via high-throughput *in situ* pairwise sequencing. *Nat. Biomed. Eng.* **8**, 872–889 (2024).

Xu, G. *et al.* The incomplete dominance of gene expression controlled by trans-eQTL hotspots contributes to heterosis in maize. *Nat. Commun.* **17**, 6551 (2026).

Xu, W. *et al.* U-FISH: a fluorescent spot detector for imaging-based spatial-omics analysis and AI-assisted FISH diagnosis. *Genome Biol.* **26**, 261 (2025).

Yu, P. *et al.* Seedling root system adaptation to water availability during maize domestication and global expansion. *Nat. Genet.* **56**, 1245–1256 (2024).

Yu, P., Eggert, K., von Wirén, N., Li, C. & Hochholdinger, F. Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. *Plant Physiol.* **169**, 690–704 (2015).

Yu, P., Gutjahr, C., Li, C. & Hochholdinger, F. Genetic control of lateral root formation in cereals. *Trends Plant Sci.* **21**, 951–961 (2016).

Zhai, R. *et al.* Transcriptome analysis of rice root heterosis by RNA-seq. *BMC Genomics* **14**, 19 (2013).

Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

Zhou, P., Hirsch, C. N., Briggs, S. P. & Springer, N. M. Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. *Mol. Plant* **12**, 410–425 (2019).

Contributions

P.Y., F.H. and X.C. conceived and designed the study. P.Y. coordinated the overall research project. P.Y. and S.J. supervised the single-cell and spatial transcriptomic data analyses. D.W. and C.W. performed all single-cell and spatial transcriptomic data analyses. Y.Z. and G.W. conducted the experimental work, including plant cultivation, root phenotyping, single-cell and spatial transcriptomic experiments, and targeted spatial expression validation. J.W. performed root protoplast isolation and sample preparation for single-cell RNA sequencing. M.P. contributed to genetic mapping and analysis of the IBM recombinant inbred line-derived population. F.H., S.J., X.C. and Y.X. contributed to data interpretation. D.W. and P.Y. drafted the manuscript. P.Y. led the integration and interpretation of the findings and revised the manuscript with input from all co-authors. All authors reviewed and approved the final manuscript.

Acknowledgements

We thank Selina Siemens (University of Bonn) for her technical assistance with sample preparation and laser-capture microdissection (LCM), Fanghe Jin for her assistance with plant cultivation and preparation of root sections, and Liang Zong (BGI Genomics) for providing technical support during root protoplast isolation and preparation for single-cell RNA sequencing.

Funding

This work is supported by the Deutsche Forschungsgemeinschaft (DFG) grants 439906817, 514003603, 552843484, the DFG Emmy Noether Programme 444755415, and the Bavarian State Ministry of Science and the Arts (bayklif2) and the German Federal Ministry of Research, Technology and Space (BMFTR) under grant number 031B1537E to P.Y., X.C.'s research is supported by The Changjiang Scholarship, Ministry of Education, China, State Cultivation Base of Eco-agriculture for Southwest Mountainous Land (Southwest University, Chongqing, China) and the National Maize Production System in China (grant no. CARS-02-15). Y.X.'D research is supported by Hong Kong Research Grants Council General Research Funding 12103126.

Competing interests

The authors declare no competing interests.

Supplementary information

This file contains Supplementary Fig. 1–24 and Supplementary Tables 1–13.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SpatialSinglecellheterosismssuppl.pdf
- SpatialSinglecellheterosismssupptable.xlsx