

# SCIENTIFIC REPORTS



OPEN

## Characterization of factors underlying the metabolic shifts in developing kernels of colored maize

Received: 15 July 2016  
Accepted: 30 September 2016  
Published: 14 October 2016

Chaoyang Hu<sup>1</sup>, Quanlin Li<sup>1</sup>, Xuefang Shen<sup>2</sup>, Sheng Quan<sup>1</sup>, Hong Lin<sup>1</sup>, Lei Duan<sup>1</sup>, Yifa Wang<sup>2</sup>, Qian Luo<sup>1</sup>, Guorun Qu<sup>1</sup>, Qing Han<sup>2</sup>, Yuan Lu<sup>2</sup>, Dabing Zhang<sup>1,3,4</sup>, Zheng Yuan<sup>1,4</sup> & Jianxin Shi<sup>1,4</sup>

Elucidation of the metabolic pathways determining pigmentation and their underlying regulatory mechanisms in maize kernels is of high importance in attempts to improve the nutritional composition of our food. In this study, we compared dynamics in the transcriptome and metabolome between colored SW93 and white SW48 by integrating RNA-Seq and non-targeted metabolomics. Our data revealed that expression of enzyme coding genes and levels of primary metabolites decreased gradually from 11 to 21 DAP, corresponding well with the physiological change of developing maize kernels from differentiation through reserve accumulation to maturation, which was cultivar independent. A remarkable up-regulation of anthocyanin and phlobaphene pathway distinguished SW93 from SW48, in which anthocyanin regulating transcriptional factors (*R1* and *C1*), enzyme encoding genes involved in both pathways and corresponding metabolic intermediates were up-regulated concurrently in SW93 but not in SW48. The shift from the shikimate pathway of primary metabolism to the flavonoid pathway of secondary metabolism, however, appears to be under posttranscriptional regulation. This study revealed the link between primary metabolism and kernel coloration, which facilitate further study to explore fundamental questions regarding the evolution of seed metabolic capabilities as well as their potential applications in maize improvement regarding both staple and functional foods.

Maize (*Zea mays* ssp. *mays* L.) is of global significance not only as a food, feed and a source of diverse industrial products, but also as a model system with tremendous genetic diversity for general plant biology studies as well as for specific biological phenomena such as heterosis and gene transposition<sup>1,2</sup>. Its storage organs, the kernels, contain essential chemical components necessary for seed germination and seedling growth, and essential nutrients for human food and livestock feed. Similar to other cereals such as rice, the value of the maize kernel is mainly determined by its endosperm<sup>3</sup>. Therefore, the development of maize endosperm has attracted special attention in the past decades. Maize kernel development includes endosperm development and embryo development, two separate but associated phases. The three endosperm development stages are distinct but overlapped. The early stage initiates from 1 to 4 DAP, containing double fertilization, syncytium formation and cellularization; the differentiation stage lasts from 4 to 20 DAP, including cell type formation, mitotic cell proliferation, endoreduplication and reserve accumulation; while the maturation stage begins from around 16 to about 60 DAP, comprising programmed cell death, dormancy and desiccation<sup>3</sup>. Along the endosperm development, the maize embryo develops from a pro-embryo (5 DAP) into a well differentiated embryo (15 DAP) in several distinguishable but continuous stages, including pro-embryo formation and initiation (7 DAP), rapid enlargement of the scutellum (9–12 DAP), and embryo maturation (13–15 DAP)<sup>4,5</sup>. The timing of kernel development, however, depends largely on genetic background and environmental conditions<sup>5</sup>.

Previous studies reveal that the physiological and morphological changes along maize kernel development result mainly from developmental cue-regulated gene expression and the functions of the proteins that they encode at specific developmental stages. At the transcriptional level, the gene expression profiles of developing

<sup>1</sup>State Key Laboratory of Hybrid Rice, Shanghai Jiao Tong University–University of Adelaide Joint Centre for Agriculture and Health, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. <sup>2</sup>Shanghai Academy of Agricultural Sciences, Shanghai 201403, China. <sup>3</sup>Plant Genomics Center, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Urrbrae, South Australia 5064, Australia. <sup>4</sup>Key Laboratory of Crop Marker-Assisted Breeding of Huaian Municipality, Jiangsu Collaborative Innovation Center of Regional Modern Agriculture and Environmental Protection, Huaian 223300, China. Correspondence and requests for materials should be addressed to Z.Y. (email: zyuan@sjtu.edu.cn) or J.S. (email: jianxin.shi@sjtu.edu.cn)

maize kernels are clustered into three distinct groups: 2–8 DAP, 10–14 DAP and 16–24 DAP<sup>6</sup>, those of developing maize endosperm are classified into 12–18 DAP and 20–24 DAP groups, while the differences of transcriptomes of maize embryos at 16–24 DAP are much smaller than those of seeds and endosperms<sup>6</sup>. Another transcriptomic study revealed remarkable changes in the pattern of certain functional genes in endosperm transcriptome during storage accumulation stage (10, 14 and 21 DAP)<sup>7</sup>. In addition, extensive proteomic studies on maize kernel development revealed that the abundance of proteins that are involved in cell division, respiration, detoxification and energy production is higher at early stages, while that of proteins related to storage protein synthesis, protein folding and glycolysis is higher at relative late stages<sup>8–11</sup>. These findings described above indicated that maize kernel development is regulated at least at both transcriptional and post-transcriptional levels.

Metabolites, the final products of the cellular regulatory process<sup>12</sup>, represent the building blocks for macromolecules and are also an essential component of cellular energy. Metabolomic analysis of the kinetic metabolic patterns of developing seeds in plants such as *Arabidopsis*<sup>13</sup>, tomato<sup>14</sup> and rice<sup>15</sup>, have provided important molecular and biochemical insight into our understanding of seed development. In maize, a pioneer study reported that the contents of dry weight, total nitrogen, protein, fat and DNA in a kernel increase gradually along seed development while those of total amino acids and soluble nucleotides first increase to the maximum value at 28 DAP and decrease afterwards<sup>16</sup>. Several recent excellent works concerning the natural variation of mature maize kernel metabolomes also provide useful information about the chemical diversity in maize kernels<sup>17–20</sup>, however, a detailed comprehensive analysis of the dynamic metabolic changes in developing kernels has not been carried out. Determination of the molecular and biochemical mechanisms governing kernel development in maize is of significance not only for basic research on maize development but also for maize improvement via metabolic engineering.

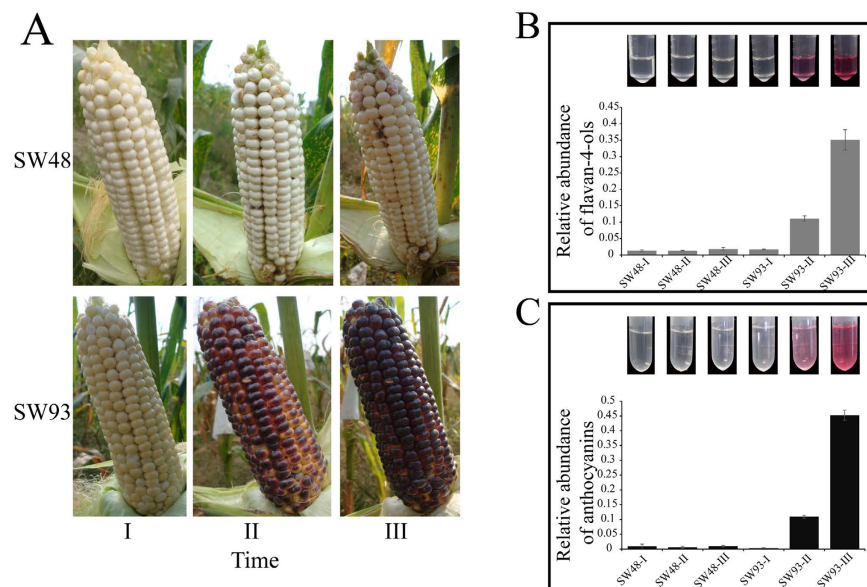
Color development is an important event of maize kernel development. During their adaptation to new growth environments and particularly due to artificial selection by humans, maize kernels lost their colors and as a consequence globally cultivated maize are mainly yellow kernels<sup>21</sup>. Recently, there is increasing demand for dark-colored maize kernels because of the rising awareness of the nutritional importance of pigments in the prevention of chronic diseases<sup>22–24</sup>. Therefore, considerable attentions have been drawn back to ancient flavonoid-rich landraces with attempts to understand the color formation in a range of crops including maize kernels<sup>21,25</sup>. It is well known that flavonoids, mainly anthocyanins and specifically phlobaphenes, account for the colorful pigments in maize kernels<sup>21,26,27</sup>. Anthocyanin compounds are responsible for blue, purple and red hues, while phlobaphenes contribute mainly to red and purple hues<sup>21,26,27</sup>. The anthocyanin pathway in maize is transcriptionally regulated by a ternary complex of transcription factors, including the MYB factor C1 (colorless 1) or PL1 (purple leaf 1), the bHLH factor R1 (red color 1) or B1 (booster 1), and the WD40 factor PAC1 (pale aleurone color 1)<sup>28–30</sup>. Although each transcription factor appears to be regulated independently, the expression of all three is required for anthocyanin pigmentation<sup>21,31</sup>. By contrast, the biosynthesis of phlobaphene is controlled by the MYB member P1 (Pericarp color1) in the absence of R1 or B1<sup>27,32,33</sup>. Transcriptionally, these transcriptional factors function via regulating three common flavonoid biosynthetic structural genes, *chalcone synthase* (*CHS*), *chalcone isomerase* (*CHI*) and *dihydroflavonol 4-reductase* (*DFR*), to produce precursors of various anthocyanins and phlobaphenes. In addition, because of the remarkably uncompact structure of the promoters of *R1*<sup>34–36</sup>, *B1*<sup>35,37</sup>, and *P1*<sup>38,39</sup> genes, epigenetics also plays important roles in the regulation of the complicated spatial and temporal pigmentation in maize. Nevertheless, most current understanding of maize kernel pigmentation has been obtained from conventional genetic studies using mutants of the above structural and regulatory genes. To date a systematic investigation of the molecular and biochemical mechanisms underlying the complex pigmentation in maize kernels is still missing.

In this study, taking the advantage of the system biology approach, we comparatively investigated the dynamic transcriptomic and metabolomic profiles of developing maize kernel in two different maize cultivars, SW48 and SW93. By interactively comparing metabolomic and transcriptomic data between SW48 and SW93, we were able to draw a comprehensive picture of events underpinning the maize kernel development corresponding well with the respective physiological and phenotypic changes. The transcriptional and metabolic changes was in close agreement with the onset of pigmentation in SW93 maize kernels, and the comparison of the two lines allowed considerable insight into the regulatory mechanisms underlying pigmentation in maize kernels

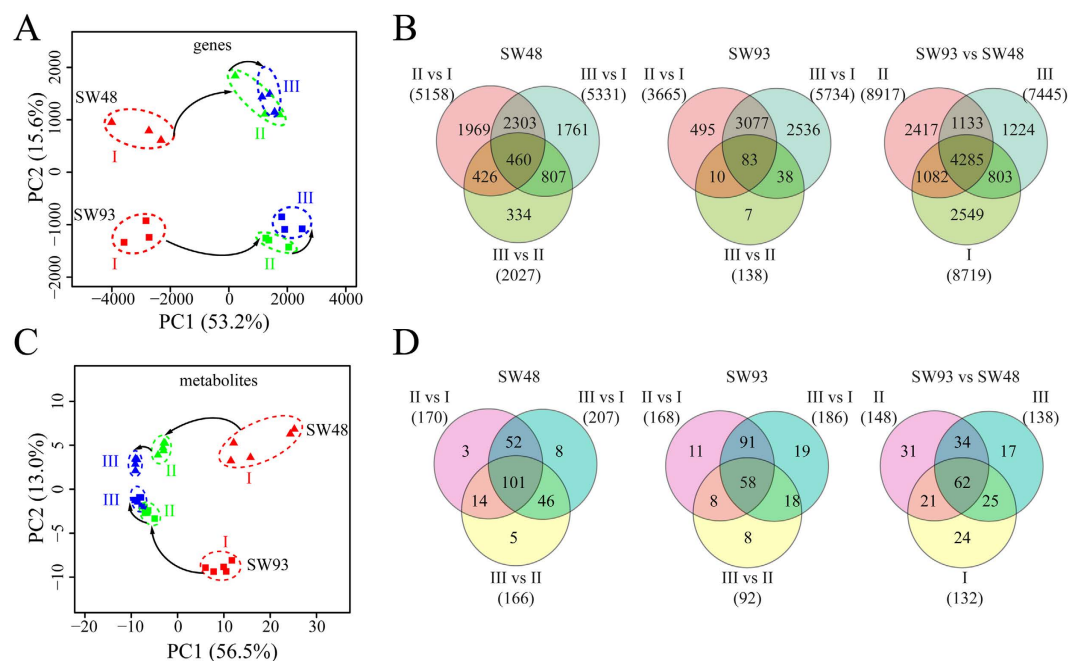
## Results

**Changes in Pigmentation Pattern.** At 11 DAP, the maize kernel of SW48 and SW93 were both white. At 16 and 21 DAP, the kernel color of SW93 changed from light purple or red to dark purple while that of SW48 remained white (Fig. 1A). Notably, across tested developmental stages, flavan-4-ols (the precursor of the phlobaphene) and total anthocyanins accumulated in SW93 but not in SW48 at 16 and 21 DAP (Fig. 1B), respectively, corresponding well with the kernel coloration patterns (Fig. 1A), indicating that two flavonoids branches (phlobaphene and anthocyanin pathways) are associated with the kernel coloration in SW93.

**Overall Kinetics of Transcriptomic Changes.** Transcriptomic analysis based on RNA-seq revealed that about 51,000 transcripts (accounting for 46% of all maize transcripts) were expressed in developing maize kernels (Supplementary Table S1). Principle component (PC) analysis on all those expressed transcripts indicated that PC 1, explaining 53.2% of the total variance, separated samples at 11 DAP from those at 16 DAP and 21 DAP (Fig. 2A), and demonstrated that the transcriptome of maize kernel at 11 DAP was significantly different from those at 16 and 21 DAP in both cultivars. The fact that samples at 16 and 21 DAP were closely grouped and almost indistinguishable in both cultivars (Fig. 2A) indicated that in both cultivars the transcriptomes of developing kernels at these two stages were highly similar. PC 2, accounting for 15.6% of the total variance, separated samples of SW48 from those of SW93 (Fig. 2A), which demonstrated distinct kernel transcriptomes between SW48 and SW93 at all three tested time points. Notably, the scores of samples at 11 DAP and those at 16 and 21 DAP were



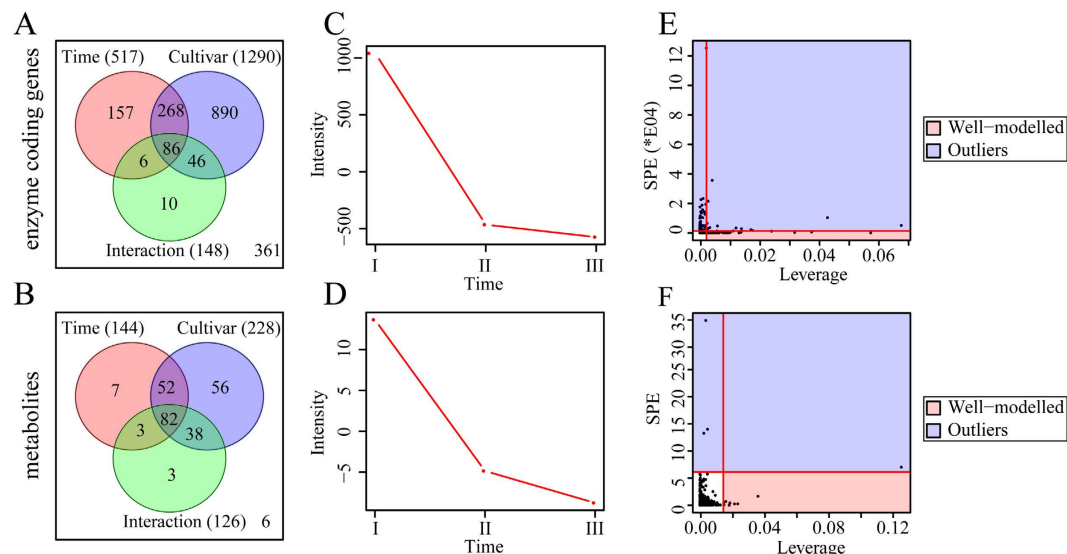
**Figure 1. The development and coloration of maize kernels.** (A) Phenotypes of developing maize ears of SW48 and SW93 along the development in the field; (B) The accumulation of phlobaphene precursors in SW93 kernels along the development; (C) The accumulation of anthocyanins in SW93 kernels along the development.



**Figure 2. The overview of the transcriptomic and metabolomic changes in developing maize kernels.**

(A,C) Principle component analysis of detected transcriptomic data and metabolomic data, respectively. Red, green and blue colors represent samples at I, II and III, respectively. Box and triangle denote samples of SW93 and SW48, respectively. I, II and III refer to 11, 16 and 21 DAP, respectively. (B,D) Venn diagrams of differentially expressed genes (DEGs), and differential metabolites (DAMs), of maize kernel at different time points of the same cultivar and at the same time point of different cultivars. I, II and III refer to 11, 16 and 21 DAP, respectively.

distributed on the left and right sides of the plot, respectively (Fig. 2A). When PCA was applied to all detected 1867 enzyme coding genes (ECG) based on Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.kegg.jp/>) and Plant Metabolic Pathway Database (<http://www.plantcyc.org/>), similar separation patterns of samples at 11 DAP from those at 16 and 21 DAP was also observed (Supplementary Fig. S1). However, the samples at 11 DAP and those at 16 and 21 DAP were distributed on the right and left sides of the plot, respectively (Supplementary Fig. S1), notably different from that of all expressed genes (Fig. 2A).



**Figure 3. Result of two-way ANOVA and ASCA. (A,B)** Venn diagram of enzyme coding genes and metabolites whose levels were affected by time, cultivar or time-cultivar interaction; **(C,D)** Major pattern of enzyme coding genes and metabolites associated with time. I, II and III refer to 11, 16 and 21 DAP, respectively; **(E,F)** ASCA selection of important enzyme coding genes and metabolites associated with time by Leverage/SPE analysis. These analysis was performed in MetaboAnalyst website <http://www.metabolanalyst.ca> website<sup>63</sup>.

In SW48, as compared with samples of 11 DAP, the numbers of differentially expressed genes (DEG) in samples of 16 DAP and 21 DAP were 5,158 (3,166 up and 1,992 down) and 5,331 (3,557 up and 1,774 down), respectively, and as compared with samples of 16 DAP, the number of DEG in samples of 21 DAP was 2,027 (1,322 up and 705 down) (Fig. 2B). Similarly, in SW93, as compared with samples of 11 DAP, the numbers of DEG in samples of 16 DAP and 21 DAP were 3,665 (2,547 up and 1,118 down) and 5,734 (3,921 up and 1,813 down), respectively, and as compared with samples of 16 DAP, the number of DEG in sample of 21 DAP was 138 (124 up and 14 down) (Fig. 2B). These data indicated that the numbers of developmental associated DEGs increased similarly in both cultivars.

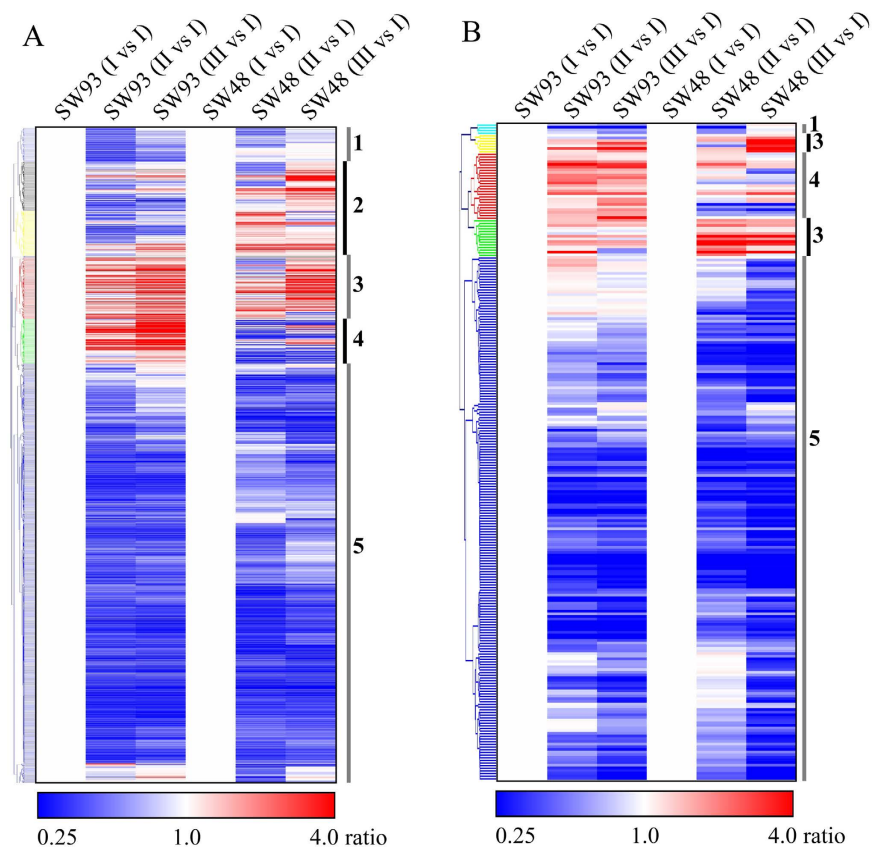
**Overall Kinetics of Metabolomic Changes.** Metabolomic analysis based on UHPLC-MS/MS and GC-MS identified a total of 247 metabolites with known structures in developing maize kernels (Supplementary Table S2). PCA on these metabolites uncovered that PC1, explaining 56.5% of total variance, separated samples at 11 DAP from those at 16 and 21 DAP, and PC2, accounting for 13.0% of left variance, separated SW93 from SW48 (Fig. 2C). Thus, metabolome of kernels at 11 DAP was significantly different from those at 16 and 21 DAP in both cultivars, and metabolomes of SW48 at 11, 16 and 21 DAP were distinct from those of SW93. Remarkably, the metabolite scores of samples at 11 DAP and those at 16 and 21 DAP were distributed on the right and left sides of the plot, respectively, which is similar to the plot of ECGs (Supplementary Fig. S1). Therefore, unless stated, gene expression analysis targets ECGs.

In SW48, as compared with samples of 11 DAP, the numbers of differentially altered metabolites (DAM) ( $P$ -value  $\leq 0.05$ ) in samples of 16 DAP and 21 DAP were 170 and 207, respectively, and as compared with samples of 16 DAP, the number of DAM in sample of 21 DAP was 116 (Fig. 2B). Similarly, in SW93, as compared with samples of 11 DAP, the numbers of DAM in samples of 16 DAP and 21 DAP were 168 and 186, respectively, and as compared with samples at 16 DAP, the number of DAM in sample of 21 DAP was 92 (Fig. 2D). This revealed metabolomic kinetics is similar to that of transcriptomic kinetics (Fig. 2B).

**Transcriptomic and Metabolomic Variations between SW48 and SW93.** There were remarkable transcriptional and metabolomic variations between SW48 and SW93 across kernel development. As compared with SW48, the numbers of DEG and DAM in SW93 samples at 11 DAP were 8,719 and 132, respectively, those at 16 DAP were 8,917 and 148, respectively, and those at 21 DAP were 7,445 and 138, respectively (Fig. 2B and Fig. 2D).

Because PCA is not able to distinguish the exact contribution of each variable to the observed variation, therefore, two-way ANOVA and ASCA were conducted to decompose both transcriptional and metabolic variations derived from time, cultivar and their interaction (Fig. 3). ANOVA revealed that the numbers of DEG affected significantly by time, cultivar and their interaction were 517, 1,290, and 148, respectively (Fig. 3A, Supplementary Table S3), while those of DAM were 144, 228 and 126, respectively (Fig. 3B, Supplementary Table S4). In addition, ASCA revealed that 55.5%, 10.0% and 14.2% of observed variations of ECG (Supplementary Fig. S2), and 55.5%, 9.7%, and 12.5% of observed variations of metabolite levels (Supplementary Fig. S3), could be explained by time, cultivar and their interaction, respectively. Time score plots based on PC1 of the corresponding submodels also showed similar transcriptional and metabolomic changes, in both cases, the scores gradually decreased from 11 DAP to 21 DAP (Fig. 3C,D), indicating a conserved and coordinated reduction in the transcriptional and





**Figure 4. Hierarchical clustering analysis of the transcriptomic and metabolomic data.** (A) enzyme coding genes; (B) metabolites. Red color indicate the abundance of enzyme coding gene or metabolite tends to be up-regulated compared to that at I of the same cultivar and blue down-regulated see ratio color key. 1 and 5, cluster 1 and cluster 5: down-regulated in both SW93 and SW48; 2, cluster 2: down-regulated in SW93, but up-regulated in SW48; 3, cluster 3: up-regulated in both SW93 and SW48; 4, cluster 4: up-regulated in SW93, but down-regulated in SW48.

metabolic activities across maize kernel development. Furthermore, Leverage/SPE analysis<sup>40,41</sup> revealed that the numbers of identified ECGs responsible for observed variations derived from time, cultivar, and their interaction were 56, 32, and 25, respectively (Fig. 3E; Supplementary Table S5), and those for DAMs were eight, nine, and seven, respectively (Fig. 3F; Supplementary Table S6). Further investigating deeply into those responsible ECGs and DAMs uncovered that each variable has specific ECGs and/or DAMs (Supplementary Tables S5 and S6). For example, the 56 ECGs identified as time variable were associated mainly with energy metabolism; the 32 ECGs identified as being cultivar variable were involved in 21 metabolic pathways; the 25 ECGs identified as being dependent on the interaction variable participated in diverse metabolic pathways including the flavonoid biosynthetic pathway (Supplementary Table S5). Similarly, the eight DAMs identified as being time variable were involved in the metabolisms of amino acids, lipids and auxin; the nine DAMs identified as being cultivar variable participated in the metabolisms of amino acids, carbohydrates, cofactors and phenylpropanoids; while the seven DAMs identified as being dependent on the interaction variable were associated with the metabolisms of cofactors, nucleotides, lipids and flavonoids (Supplementary Table S6).

### Conserved and Divergent Developmental Induced Transcriptomic and Metabolomic Alterations between SW48 and SW93.

GO analysis on DEGs revealed that these two cultivars differed in GO terms associated with signaling, metabolic, and stress responsive activities at 11 DAP, in starch metabolic, stress responsive and oxidative-reductive activities at 16 DAP, and in metabolic and stress responsive activities at 21 DAP (Supplementary Table S7). Notably, GO terms related to metabolism, such as “anthocyanin-containing compound biosynthetic process” and “flavonoid biosynthetic process” were markedly enriched at 21 DAP in SW93.

Hierarchical clustering analysis on 1867 differentially expressed ECGs showed that 72% DCGs (1,348/1,867) were down-regulated at 16 and/or 21 DAP as compared with those at 11 DAP in both cultivars, although the degree of the decrease varied among genes (Fig. 4A, cluster 1). This result is consistent with the ASCA time score plot (Fig. 3C) and these downregulated genes covered almost all the detected metabolic pathways (Supplementary Table S8). The expression of 246 ECGs were upregulated in SW48 but downregulated or unaltered in SW93 at 16 and/or 21 DAP (Fig. 4A, cluster 2) and 80% of them were involved in primary metabolism while the other 20% participated in secondary metabolism (Supplementary Table S8). The expressions of 170 ECGs were upregulated in both SW93 and SW48 at 16 and/or 21 DAP (Fig. 4A, cluster 3), all these ECGs were involved in either primary

or secondary metabolism. The expression of 91 ECGs were upregulated in SW93 but were downregulated or not significantly changed in SW48 at 16 and/or 21 DAP (Fig. 4A, cluster 4), ECGs involved in flavonoids metabolism were found among them (cluster 4).

Hierarchical clustering analysis on identified metabolites showed that abundances of 81.4% (201/247) of the identified metabolites decreased at 16 and/or 21 DAP as compared with those at 11 DAP in both kernels (Fig. 4B, cluster 1), which was similar to the time score plot from ASCA (Fig. 3D). The levels of 21 metabolites increased at 16 and/or 21 DAP in both SW93 and SW48, but with relative higher levels in SW48 than those in SW93 (Fig. 4B, cluster 3); these metabolites were involved mainly in primary metabolism (Supplementary Table S9). The levels of 11 metabolites increased at 16 and 21 DAP significantly in SW93 while slightly changed at 16 and/or 21 in SW48 (Fig. 4B, cluster 4); among them, naringenin is a secondary metabolite.

**The Flavonoid Pathway Is Activated in Developing SW93 Kernels.** Increases in the contents of total anthocyanins and flavan-4-ol in SW93 kernels at 16 and 21 DAP (Fig. 1) led us to focus on the transcriptional and metabolic changes in flavonoid pathway (Fig. 5). At the transcriptional level, RNA-Seq data found that neither anthocyanin regulator *B1* (GRMZM2G172795), *PL1* (GRMZM2G701063) nor phlobaphene regulator *P1* (GRMZM2G084799 and GRMZM2G057027) was detected in kernels of SW93 and SW48 across kernel development (Supplementary Table S1). However, the expression levels of anthocyanin regulators *R1* (GRMZM5G822829) and *C1* (GRMZM2G005066) were remarkably increased in SW93 but not in SW48 across kernel development (with the exception of *C1* at 21 DAP) (Fig. 5D). The expression level of the third anthocyanin regulator, a WD40 protein, *PAC1* (GRMZM2G058432), was expressed highly in both kernels at 11 DAP, which declined similarly at 16 and 21 DAP. Notably, two other transcription factors, *AC194965.4\_FG004* (GATA family) and *GRMZM2G065374* (bHLH family), showed similar expression patterns to *R1* in SW93 (Supplementary Fig. S4). In addition, the expression levels of two structural genes commonly involved in both anthocyanin and phlobaphene pathways namely *CHS* and *DFR* were upregulated in SW93 across kernel development, including two *CHS* (GRMZM2G151227 and GRMZM2G422750) and two *DFR* (GRMZM2G013726 and GRMZM2G026930) (Fig. 5D). Notably, one *CHI* transcript (GRMZM2G119186) increased across kernel development in both kernels (Fig. 5D). Furthermore, the levels of transcripts of *ANS* (anthocyanidin synthase) (GRMZM2G345717), *F3H* (flavanone 3-hydroxylase) (GRMZM2G062396), and *UFGT* (UDP-flavonoid glucosyl transferase) (GRMZM2G165390), all displayed similar expression patterns as *CHS* and *DFR* across kernel development. These findings indicated a concurrent activation of both anthocyanin and phlobaphene pathways in SW93. The activation of phlobaphene pathway in SW93 was also evidenced by the increased expression of one transcript of *F3'H* (flavanone 3'-hydroxylase) (GRMZM2G025832), a direct target of *P1*<sup>42</sup>, in SW93 but not in SW48 (Fig. 5D). Moreover, RNA-Seq data demonstrated that although the expression levels of transcripts of upstream genes of seven *PAL* (phenylalanine ammonia-lyase), three *C4H* (cinnamate 4-hydroxylase), and eight *4CL* (cinamic acid 4-hydroxylase) fluctuated across kernel development in SW93, their patterns were quite similar to those in SW48 (Fig. 5D). This divergence was also confirmed by the observed differential expression levels of transcripts of genes downstream of *FLS* (flavonol synthase) (Fig. 5D). The expression pattern changes of several genes across kernel development from 11 to 21 DAP were confirmed by qualitative RT-PCR (Fig. 5C).

Although non-targeted UPLC-MS-MS only identified three metabolites in the phenylalanine pathway, they all exhibited distinct kinetic patterns across kernel development between SW93 and SW48. The abundance of phenylalanine declined continuously in SW48 whilst it was unaltered in SW93 across kernel development, the level of coumaric acid in SW93 was always significantly higher than that in SW48 across kernel development, while the level of naringenin, was remarkably increased across kernel development in SW93 whilst being kept at a basal level in SW48 (Fig. 5B). Altogether, these findings strongly indicated the developmental clue redirected metabolic shift to flavonoids pathway in SW93, leading to colored kernels.

**Other Genotype Specific Changes.** *Glycolysis.* RNA-Seq data revealed that the expression levels of most of the 54 genes encoding 12 enzymes involved in glycolysis were down-regulated across kernel development in both kernels (Supplementary Fig. S5B), while those of transcripts coding for a hexokinase (EC 2.7.1.1), a 6-phosphofructokinase (EC 2.7.1.11) and a fructose-bisphosphate aldolase (EC 4.1.2.13) were significantly up-regulated only in SW93 at 16 and 21 DAP (Supplementary Fig. S5B). On the other hand, metabolomic analysis showed that levels of three glycolytic intermediates (glucose, glucose-6P and fructose-6P) decreased and that of phosphoenolpyruvate (PEP) increased in both cultivars during development (Supplementary Fig. S5A). Notably, different from that in SW48, the level of pyruvate, the end-product of glycolysis and the indispensable precursor of the tricarboxylic acid (TCA) cycle remained constant across kernel development in SW93 (Supplementary Fig. S5A).

*The TCA cycle.* RNA-Seq data revealed that the expression levels of almost all TCA genes were significantly down-regulated at 16 and 21 DAP in both cultivars (Supplementary Fig. S6C). Notably, those of the transcripts coding for citrate oxaloacetate-lyase (EC 2.3.3.1) and for isocitrate dehydrogenase (EC 1.1.1.42) were significantly up-regulated in SW93 (Supplementary Fig. S6C). Six metabolites involved in TCA cycle were identified and the levels of them all decreased in both SW93 and SW48 (Supplementary Fig. S6A). It is worth noting however that the level of  $\alpha$ -ketoglutarate in SW93 was consistently higher than that in SW48 across kernel development, and that of citrate at 21 DAP in SW93 was obviously higher than that in SW48, which is in close agreement with above-mentioned upregulation of the genes encoding for citrate oxaloacetate lyase and isocitrate dehydrogenase.

*Other differentially altered pathways.* Other pathways were differentially affected between the genotypes but these changes were mainly apparent at the metabolite level. In pentose phosphate pathway, the level of glucose, the precursor of this pathway was stable, while the levels of the two pentose intermediates xylulose-5-phosphate and ribose-5-phosphate, were lower in SW93 kernel across kernel development (Supplementary Fig. S7). In purine



## Discussion

During seed development, various metabolites are synthesized and stored, making seeds chemical factories of stored nutrients. Understanding the complete and holistic knowledge of seed development events, particularly their reserve accumulation mechanisms, is therefore a primary prerequisite for guaranteeing the global food security via knowledge-based genetic improvement<sup>43</sup>. In this study, we comparatively characterized the dynamic transcriptome and metabolome of two maize kernels of different pigmentation during the reserve accumulation and maturation phases of seed development. Our results revealed many conserved transcriptomic and metabolomic events in those two maize cultivars, which are remarkably similar to reported observations in other dicot plants<sup>13,14</sup>. However, our data also uncovered a considerable molecular (transcriptional and metabolic) shift in colored maize kernels; potential regulatory aspects uncovered by this are discussed below in the context of the flavonoids biosynthesis.

**Transcriptomic and Metabolomic Kinetics of Maize Kernel Development.** Based on both transcriptomic and metabolomic data obtained in this study, there was clear decline in gene expression and enzyme metabolic activity in maize kernels from 11 DAP till 21 DAP, corresponding to the maize seed developmental stages from reserve accumulation to maturation (Fig. 3C,D). This resulted in two developmental stage dependent transcriptomes and metabolomes, with those at 16 DAP being easily distinguishable from those at 11 DAP but only moderately distinguishable from those at 21 DAP (Figs 2 and 4). These molecular changes correspond well to the physiological and morphological changes occurring in parallel across kernel development. Kernels at 11 DAP were active in reserve accumulation, while those at 16 DAP were approaching maturation<sup>3</sup>, therefore, the active genes and resulting metabolites showed substantial decline when kernels developed from 11 DAP to 16 DAP and afterwards. The cultivar independent transcriptomic and metabolomic kinetics observed in both developing maize kernels suggested a highly conserved regulatory mechanism within maize seed development, which is quite similar to previous reports in rice<sup>15</sup>, tomato<sup>14,44</sup>, and Arabidopsis<sup>13</sup>. Notably, the genome wide expression patterns of all genes were somewhat divergent than those of ECGs (Fig. 2A and Supplementary Fig. S1), which might, at least partially, explain the imperfect match of metabolomic data to transcriptomic data<sup>45</sup> when all detected DEGs as opposed to differentially expressed ECGs are compared. Previous proteomic and transcriptomic studies have already stressed the remarkable changes in maize kernels at reserve accumulation stages<sup>6,7</sup>. Thus, our findings appear to be conserved across a wide range of maize cultivars as well as in other crops, highlighting the value in identifying key genes for seed development even in those species whose seeds exhibit remarkable differences in size and storage components.

**The Metabolic Shifts in Developing Colored Maize Kernels.** Understanding of the pathways of pigmentation and their regulation in kernels is of high commercial importance in maize and other related cereal crops. Considerable progresses has been made in defining and understanding the metabolic pathways leading to the colorful maize kernels, however, few studies have attempted to integrate changes in primary metabolism to coloration and as such our ability to influence maize kernel coloration is arguably not as great as it should be. While this study reveals that although the general kinetic trends in transcriptome and metabolome of both SW93 and SW48 were quite similar across kernel development, the transcriptional and metabolic changes in SW93 were clearly associated with the coloration in its kernels. The substantial transcriptional and metabolic shifts in SW93 occurred evidently at 16 DAP (Fig. 4), which coincided with the appearance of pigment in SW93 kernels at this stage (Fig. 1), indicating an irreversible shift to secondary metabolism in SW93, producing anthocyanins and precursors of phlobaphene, two important secondary metabolites, to color the maize kernels with colorful pigments. This result may also be explained by a likely existence of a transcription factor in SW93 that targets both primary and secondary metabolism as reported for tomato SLMYB12<sup>24</sup> and its maize counterpart P1<sup>27</sup>. Under certain conditions, adequate supplies of precursors, energy and reducing power from the primary metabolism are required to meet demands of specialized secondary metabolism.

In plants, the shikimate pathway plays important roles in plant growth, development and response to environment by providing three aromatic amino acids (tryptophan, tyrosine and phenylalanine) that are essential precursors for the secondary metabolism, with phenylalanine being the direct precursor for pigmentation in most plants<sup>46</sup>. Whilst no clear transcriptional activation of the whole shikimate pathway was observed in developing SW93 kernels, the level of a transcript (GRMZM2G125923) encoding an arogenate/prephenate dehydratase (EC 4.2.1.91) was only upregulated in SW93, which led to a constant phenylalanine level in SW93 cross kernel development, contrasting to the significant decline observed in SW48 (Supplementary Fig. S11). The lack of a coordinated response in the shikimate pathway is arguably unsurprising since plants, unlike the AroM complex of fungi in which a single penta-functional protein is responsible for shikimate biosynthesis, do not have a gene cluster for the shikimate pathway and indeed the gene expression patterns of the constituent enzymes are somewhat distinctive<sup>47</sup>. Interestingly, tryptophan displayed similar metabolic kinetics to those of phenylalanine without upregulation of corresponding biosynthetic genes such as anthranilate synthase (EC 4.1.3.27) and chorismate mutase (EC 5.4.99.5), suggesting that stable metabolic supplies for tryptophan and phenylalanine are likely maintained by shikimate pathway via as yet unknown post-transcriptional mechanisms.

The stable supply of phenylalanine in developing kernels of SW93 was likely derived directly from glycolysis and TCA cycle. As compared with SW48, the level of PEP was constantly lower in SW93. PEP is the important entry metabolite from glycolysis to shikimate pathway, both oxaloacetate from TCA and pyruvate from glycolysis are important precursors for PEP<sup>48</sup>. Notably, levels of oxaloacetate and pyruvate in SW93 were constant but those in SW48 kept decline along the kernel development. These results indicated a metabolic shift from central carbon metabolism to amino acid metabolism in developing SW93 kernels, which could also be evidenced by the observed higher levels of amino acids and their derivatives derived from PEP (tyramine), oxaloacetate (*allo*-threonine, isoleucine and 2-aminoadipate), and pyruvate (3-methyl-2-oxovalerate) in SW93 kernels. In



addition, the higher  $\alpha$ -ketoglutarate and lower levels of  $\alpha$ -ketoglutarate derived amino acids and/or derivatives (glutamate, arginine and *N*-monomethylarginine) indicated that  $\alpha$ -ketoglutarate was not the entry metabolite from carbon flux to nitrogen flux in SW93 kernels. The higher  $\alpha$ -ketoglutarate in SW93 kernels reflected actually the consequence of the post-transcriptionally activated phenylalanine pathway via chorismate mutase, prephenate dehydratase, and phenylpyruvate aminotransferase. This metabolic shift was also observed in the coloration process of developing wild type tomato fruits<sup>49,50</sup>. This similarity suggests that it is interesting to investigate a broader range of plant systems in order to understand better the interface between primary and secondary metabolism and its impact on coloration.

**The Regulatory Mechanisms Underlying the Pigmentation in SW93 Kernel.** The majority of maize varieties carry all the biosynthetic genes of pigmentation and the difference between “colored” and “uncolored” cultivars depends on the presence of “strong alleles” of the dominant regulatory genes that up-regulate the corresponding biosynthetic genes and determine the pigmentation of specific tissues<sup>21</sup>. As reviewed recently, these transcription factors usually function via the formation of various complexes<sup>51</sup>. Our transcriptomic data obtained here demonstrated that although all biosynthetic genes involved in anthocyanin and phlobaphene pathways were detected in both cultivars, the developmental stage dependent activation of these genes was only observed in SW93 (Fig. 5). Given that the development dependent activation of the flavonoid pathway was coincident only with the activation of two anthocyanin regulators (R1 and C1), the accumulation of anthocyanin in SW93 kernels seems highly likely to be the consequence of the activation of co-existing alleles of R1 and C1, two known developmental stimuli responsive anthocyanin transcription factors<sup>21,28</sup>. However, the kinetics of the third anthocyanin regulator PAC1 along the kernel development was quite similar in both cultivars, therefore, the exact details of the functional complex among C1, R1 and PAC1 remains unknown.

The formation of phlobaphene pigments in maize is developmentally controlled by the allelic variance in P1, which results in four phenotypes, P1-rr (red pericarp and red cob), P1-wr (white pericarp and red cob) P1-rw (red pericarp and white cob), and P1-ww (white pericarp and white cob)<sup>27,52</sup>. In a recent report, the coloration of an ancient Italy maize “Nero Spinoso” (P1-rw) was characterized to be under the control of a monogenic dominant gene, a strong allele of P1. Although SW93 looked phenotypically similar to P1-rw (Fig. 1A), the genetic evidence (Supplementary Fig. S1) indicated that the pigmentation was not under the control of a monogenic dominant gene. This fact aside, it was highly surprising that the expression of P1 was not detectable in the developing kernels of either cultivars, rendering the regulation of phlobaphene pigmentation in SW93 rather puzzling. One possible explanation is that instead of P1 other regulators of phlobaphene pigmentation exist in SW93 kernels, which are developmentally regulated in a manner similar to that of R1. Our qRT-PCR result excluded the possibility of P2 as such an alternative regulator, because the expression of neither P1 nor P2 was detected (Fig. 5C,D). Perhaps the two transcription factors showing similar expression patterns to that of R1 across kernel development are possible candidate regulators, however, the conclusive identification of alternative regulators rests on the further characterization of them (Supplementary Fig. S4). An additional explanation is the likely epigenetic regulation of P1 as reported previously<sup>38</sup>, which merits further investigations into the possible existence of tissue-specific regulatory elements within the P1 non-coding sequences or flanking regions, as well as DNA methylation. Given that we could not detect P1 expression across kernel development, we favor the first explanation. Nevertheless, the regulatory mechanisms underlying the pigmentation in SW93 maize kernel provided us a good chance to explore the molecular aspects of pigmentation in developing kernels using systems biology approaches.

In summary, taking the advantages of systems biology approach, we were able to unravel the transcriptomic and metabolomic kinetics of maize kernels across the development in general and characterize factors underlying the metabolic shifts in developing kernels of colored maize in particular. The coordinated metabolic shift between primary and secondary metabolism identified in colored SW93 maize will facilitate the further efforts to understand the complex relationships between crops rich in pigments, foods, and human health, and to improve the quality of crop-based foods via metabolic engineering.

## Methods

**Maize Materials.** The two maize cultivars, SW48 (white kernels and white cobs) and SW93 (purple kernels and white cobs) were planted in the experimental fields of Shanghai Academy of Agricultural Science (Zhuanghang, Shanghai) during the summer season in 2014. These two inbred lines were used as male and female parent, respectively, to produce the popular fresh consumed sticky hybrid maize Huwucaihuanuo 1 (Supplementary Fig. S12). Maize kernels were collected at 11, 16 and 21 DAP, immediately frozen with liquid nitrogen, grounded into fine power in liquid nitrogen, aliquoted and stored at  $-80^{\circ}\text{C}$  prior to transcriptomic and metabolome analyses. Kernels taken from an individual corn cob were considered as one biological replicate, and both transcriptomic and metabolome analyses were performed with three replicates.

**Analysis of Total Anthocyanin Content.** Anthocyanins were extracted as described previously by Žilić *et al.*<sup>26</sup> with slight modifications. 120 mg fresh seed powder was extracted by mixing with 1 ml of methanol acidified with 1N HCl (85:15, v/v) and shaking for 30 min at room temperature. The crude extract was centrifuged at 12,000 rpm for 10 min, and 1 ml of the supernatant was mixed with 1 ml extraction solution. Then, the absorbance (A) of samples were measured at 535 and 700 nm, respectively, and concentrations (relative value) were calculated using the following equation:  $A = (A_{535} - A_{700}/3)^{23}$ . The assay was performed in triplicate.

**Analysis of Flavan-4-ol Content.** Flavan-4-ols were extracted as previously described by Robbins *et al.*<sup>53</sup> with minor modifications. 150 mg fresh seed powder was incubated with 500  $\mu\text{l}$  of butanol acidified with concentrated HCl (70:30, v/v) at  $37^{\circ}\text{C}$  for 1 hour. The crude extract was centrifuged at 12,000 rpm for 10 min, and the

supernatant was mixed with 1.5 ml extraction solution. The absorbance of samples were recorded at 550 nm and the relative abundance of flavan-4-ols was represented with  $A_{550}$ . The assay was performed in triplicate.

**RNA Extraction, cDNA Library Construction and RNA Sequencing.** Total RNA of each sample was extracted with the Qiagen Rneasy Mini Kit and digested with Dnase I according to the manufacturer's instructions (Qiagen, Hilden, Germany). The integrity and size distribution of the RNA were verified with Agilent 2200 Bioanalyser (Agilent Technologies, USA.). Samples with the RNA Integrity Number  $\geq 8.0$  were used for cDNA library construction. The cDNA libraries for single-ending sequencing were constructed using Ion Total RNA-Seq Kit v2.0 (Life Technologies, Thermo Fisher Scientific, USA) according to the manufacturer's instructions. The cDNA libraries were then processed for the Proton Sequencing process according to the commercially available protocols. Samples were diluted and mixed, and the mixture was processed on a OneTouch 2 instrument (Life Technologies) and enriched on a OneTouch 2 ES station (Life Technologies) for preparing the template-positive Ion PI™ Ion Sphere™ Particles (Life Technologies) according to Ion PI™ Template OT2 200 Kit v2.0 (Life Technologies). After enrichment, the mixed template-positive Ion PI™ Ion Sphere™ Particles of samples was loaded on to 1 P1v2 Proton Chip (Life Technologies) and sequenced on Proton Sequencers according to Ion PI Sequencing 200 Kit v2.0 (Life Technologies).

**RNA Sequencing Mapping.** Before read mapping, clean reads were obtained from the raw reads by removing the adaptor sequences, reads with more than 5% ambiguous bases (noted as N), and low-quality reads containing more than 20 percent of bases with qualities less than 13. The clean reads were then aligned to maize genome (version: B73) using the MapSplice program (v2.1.8). In alignment, preliminary experiments were performed to optimize the alignment parameters (*-s 22 -p 12 -ins 6 -del 6 -non-canonical*) to provide the largest information on the AS events<sup>54</sup>.

**Metabolite Profiling.** Methanol extracts from 40 mg maize kernel powder per sample were analyzed with a metabolomic platform that combines UHPLC-MS/MS with GC-MS<sup>55–57</sup>. Metabolite identification was performed as reported previously<sup>20,58</sup> based on the comparison to an in-house library entries of purified metabolite standards, containing the retention index (RI), mass to charge ratio (m/z), and MS/MS spectral data.

**Statistical Analysis.** Differentially expressed genes were filtered by applying DEseq algorithm<sup>59</sup> with Fold Change  $>2$  or  $<0.5$ , and false discovery rate (FDR)  $<0.05$ <sup>60</sup>. Gene ontology (GO) analysis was performed to facilitate elucidating the biological implications of unique genes in the significant or representative profiles of the differentially expressed gene in the experiment<sup>61</sup>. GO annotations were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>), UniProt (<http://www.uniprot.org/>) and the Gene Ontology (<http://www.geneontology.org/>). Fisher's exact test was applied to identify the significant GO categories and pathway with FDR  $<0.05$ <sup>62</sup>. Metabolite abundances were quantified using peak areas. The missing values of a giving metabolite were inputted with the detected minimum value for statistical analysis, assuming that they were below the limits of instrument detection sensitivity. Metabolic differences between different groups were compared by using Welch's two-sample t-test and significant level was declared for p-values less than 0.05.

Principle Component Analysis was performed with SIMCA-P software and "Par" was selected for data scaling type. Two-way ANOVA and ASCA (ANOVA-Simultaneous Component Analysis) were performed with MetATT<sup>63</sup> (<http://www.metaboanalyst.ca/>) using "Pareto Scaling" for data normalization. Two-way ANOVA type used is "within subjects ANOVA", significance threshold is defined as the corrected p-value  $<0.05$  and False Discovery Rate  $<0.05$  was chosen for multiple testing correction. ASCA was performed with default parameters supplied by the website. Hierarchical clustering of metabolites and enzyme coding genes (ECG) was performed with MultiExperiment Viewer (MeV) version 4.8.

## References

- Bennetzen, J. L. & Hake, S. *Handbook of maize*. (Springer, 2009).
- Serna-Saldivar, S. O., Gomez, M. H., Rooney, L. W. & Hallauer, A. Food uses of regular and specialty corns and their dry-milled fractions. *Specialty Corns*. 263–298 (1994).
- Sabelli, P. A. & Larkins, B. A. The development of endosperm in grasses. *Plant Physiol.* **149**, 14–26 (2009).
- Lammeren, A. V. Developmental morphology and cytology of the young maize embryo (*Zea mays L.*). *Acta. Botanica Neerlandica* **35**, 169–188 (1986).
- Vernoud, V., Hajdúch, M., Khaled, A., Depège, N. & Rogowsky, M. Maize embryogenesis. *Maydica* **50**, 469 (2005).
- Sekhon, R. S. *et al.* Genome-wide atlas of transcription during maize development. *Plant J.* **66**, 553–563 (2011).
- Prioul, J. L. *et al.* A joint transcriptomic, proteomic and metabolic analysis of maize endosperm development and starch filling. *Plant Biotechnol. J.* **6**, 855–869 (2008).
- Méchin, V., Thévenot, C., Le Guilloux, M., Prioul, J. L. & Damerval, C. Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiol.* **143**, 1203–1219 (2007).
- Pechanova, O., Takáč, T., Šamaj, J. & Pechan, T. Maize proteomics: an insight into the biology of an important cereal crop. *Proteomics* **13**, 637–662 (2013).
- Takáč, T., Pechan, T. & Šamaj, J. Differential proteomics of plant development. *J. proteomics* **74**, 577–588 (2011).
- Haynes, P. A. & Roberts, T. H. Subcellular shotgun proteomics in plants: looking beyond the usual suspects. *Proteomics* **7**, 2963–2975 (2007).
- Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171 (2002).
- Fait, A. *et al.* Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* **142**, 839–854 (2006).
- Toubiana, D. *et al.* Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *PLoS Genet.* **8**, e1002612 (2012).
- Hu, C. *et al.* Identification of conserved and diverse metabolic shifts during rice grain development. *Sci. Rep.* **6**, 20942 (2016).
- Ingle, J., Beitz, D. & Hageman, R. Changes in composition during development and maturation of maize seeds. *Plant Physiol.* **40**, 835 (1965).

17. Wen, W., Brotman, Y., Willmitzer, L., Yan, J. & Fernie, A. R. Broadening Our Portfolio in the Genetic Improvement of Maize Chemical Composition. *Trends Genet.* **32**, 459–469 (2016).
18. Röhlrig, R. M., Eder, J. & Engel, K.-H. Metabolite profiling of maize grain: differentiation due to genetics and environment. *Metabolomics* **5**, 459–477 (2009).
19. Skogerson, K. *et al.* Impact of genetics and environment on the metabolite composition of maize grain. *J. Agr. Food Chem.* **58**, 3600–3610 (2010).
20. Rao, J. *et al.* Metabolic map of mature maize kernels. *Metabolomics* **10**, 775–787 (2014).
21. Petroni, K., Pilu, R. & Tonelli, C. Anthocyanins in corn: a wealth of genes for human health. *Planta* **240**, 901–911 (2014).
22. Jiang, N., Doseff, A. I. & Grotewold, E. Flavones: From Biosynthesis to Health Benefits. *Plants* **5**, 27 (2016).
23. Tohge, T. *et al.* Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant J.* **83**, 686–704 (2015).
24. Zhang, Y. *et al.* Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* **6**, 8635 (2015).
25. Casas, M. I., Duarte, S., Doseff, A. I. & Grotewold, E. Flavone-rich maize: an opportunity to improve the nutritional value of an important commodity crop. *Front Plant Sci.* **5**, 1–11 (2014).
26. Žilić, S., Serpen, A., Akillioğlu, G. L., Gökmen, V. & Vančetović & J. Phenolic compounds, carotenoids, anthocyanins, and antioxidant capacity of colored maize (*Zea mays* L.) kernels. *J. Agr. Food Chem.* **60**, 1224–1231 (2012).
27. Morohashi, K. *et al.* A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *Plant Cell* **24**, 2745–2764 (2012).
28. Chandler, V. L., Radicella, J. P., Robbins, T. P., Chen, J. & Turks, D. Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. *Plant cell* **1**, 1175–1183 (1989).
29. Pilu, R. *et al.* pl-bol3, a complex allele of the anthocyanin regulatory pl1 locus that arose in a naturally occurring maize population. *Plant J.* **36**, 510–521 (2003).
30. Selinger, D. A. & Chandler, V. L. A mutation in the pale aleurone color1 gene identifies a novel regulator of the maize anthocyanin pathway. *Plant Cell* **11**, 5–14 (1999).
31. Zhao, L. *et al.* The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Funct. Integr. Genomic.* **13**, 75–98 (2013).
32. Grotewold, E. *et al.* Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. *Plant Cell* **10**, 721–740 (1998).
33. Sharma, M. *et al.* Expression of flavonoid 3'-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. *BMC Plant Biol.* **12**, 196 (2012).
34. Ludwig, S. R., Habera, L. F., Dellaporta, S. L. & Wessler, S. R. Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. *Proc. Natl. Acad. Sci. USA* **86**, 7092–7096 (1989).
35. Radicella, J. P., Brown, D., Tolar, L. & Chandler, V. Allelic diversity of the maize B regulatory gene: different leader and promoter sequences of two B alleles determine distinct tissue specificities of anthocyanin production. *Gene. Dev.* **6**, 2152–2164 (1992).
36. Li, Y. *et al.* Gene conversion within regulatory sequences generates maize r alleles with altered gene expression. *Genetics* **159**, 1727–1740 (2001).
37. Stam, M. *et al.* The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. *Genetics* **162**, 917–930 (2002).
38. Cociolone, S. M., Chopra, S., Flint-Garcia, S. A., McMullen, M. D. & Peterson, T. Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. *Plant J.* **27**, 467–478 (2001).
39. Zhang, F. & Peterson, T. Comparisons of maize pericarp color1 alleles reveal paralogous gene recombination and an organ-specific enhancer region. *Plant Cell* **17**, 903–914 (2005).
40. Nueda, M. J. *et al.* Discovering gene expression patterns in time course microarray experiments by ANOVA-SCA. *Bioinformatics* **23**, 1792–1800 (2007).
41. Cordero, F., Botta, M. & Calogero, R. A. Microarray data analysis and mining approaches. *Brief. Funct. Genomics* **6**, 265–281 (2007).
42. Sharma, R. *et al.* Expression dynamics of metabolic and regulatory components across stages of panicle and seed development in indica rice. *Funct. Integr. Genomic.* **12**, 229–248 (2012).
43. Sreenivasulu, N. & Wobus, U. Seed-development programs: a systems biology-based comparison between dicots and monocots. *Plant Biol.* **64**, 189–217 (2013).
44. Mounet, F. *et al.* Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics* **3**, 273–288 (2007).
45. Fernie, A. R. & Stitt, M. On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions scientific correspondence. *Plant Physiol.* **158**, 1139–1145 (2012).
46. Maeda, H. & Dudareva, N. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* **63**, 73–105 (2012).
47. Tohge, T., Watanabe, M., Hoefgen, R. & Fernie, A. R. Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* **4**, 62 (2013).
48. Bonner, J. & Varner, J. E. *Plant biochemistry*. (Elsevier, 2012).
49. Carrari, F. *et al.* Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.* **142**, 1380–1396 (2006).
50. Mintz-Oron, S. *et al.* Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol.* **147**, 823–851 (2008).
51. Xu, W., Dubos, C. & Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **20**, 176–185 (2015).
52. Cassani, E. *et al.* Genetic studies regarding the control of seed pigmentation of an ancient European pointed maize (*Zea mays* L.) rich in phlobaphenes: the “Nero Spinoso” from the Camonica valley. *Genet. Resour. Crop Ev.* 1–13 (2016).
53. Robbins, M. L. *et al.* Comparative proteomics analysis by DIGE and iTRAQ provides insight into the regulation of phenylpropanoids in maize. *J. Proteomics* **93**, 254–275 (2013).
54. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38**, 178 (2010).
55. Ohta, T. *et al.* Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol. Pathol.* **37**, 521–535 (2009).
56. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).
57. Clarke, J. D. *et al.* Assessment of genetically modified soybean in relation to natural variation in the soybean seed metabolome. *Sci. Rep.* **3**, 3082 (2013).
58. Hu, C. *et al.* Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.* **4**, 5067 (2014).
59. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, 106 (2010).

60. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
61. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
62. Draghici, S. *et al.* A systems biology approach for pathway level analysis. *Genome Res.* **17**, 1537–1545 (2007).
63. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res.* **43**, 251–257 (2015).

### Acknowledgements

This work was supported by grants from the Project on Breeding from Agriculture Commission of Shanghai (Grant Nos. 2013–13, 2014-1/3 and 2015-20), Key Project on Basic Research from Science and Technology Commission of Shanghai (Nos. 14JC1403900 and 14391917100), China National Transgenic Plant Special Fund (2016ZX08012-002), China Innovative Research Team, Ministry of Education, 111 Project (B14016). Authors thank Dr. Alisdair Fernie and Dr. Takayuki Tohge from the Molecular Plant Physiology Institute, the Max Planck, Germany, for their critical reading of the manuscript. We declare that there is no conflict of interest.

### Author Contributions

C.H. and Q.L. conducted most of the experiments, Q.H. and Y.L. performed part of the experiment, D.Z., Z.Y. and J.S. designed the experiments, Z.Y. and J.S. supervised the experiments, X.S. and Y.W. provided the maize material and assisted in physiological analysis, S.Q., Q.L. and G.Q. helped the metabolomic analysis, H.L. and L.D. helped data analysis. C.H., Q.L., Z.Y. and J.S. analyzed data and wrote the paper.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Hu, C. *et al.* Characterization of factors underlying the metabolic shifts in developing kernels of colored maize. *Sci. Rep.* **6**, 35479; doi: 10.1038/srep35479 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016