A distant pathway-cross regulation designed with transcription factors and JAZ repressors downregulates the biosynthesis of tobacco alkaloids and specific nitrosamines

Mingzhuo Li[#], Xianzhi He[#], Christophe La Hovary[#], Yue Zhu, Yilun Dong, Shibiao Liu, Hucheng Xing, Yajun Liu, Yucheng Jie, Dongming Ma, Seyit Yuzuak, and De-Yu Xie*

Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA

These three authors are co-first authors.

* Corresponding author, dxie@ncsu.edu

Running title: **Engineering of Distant Pathway-Cross Regulation**

Abstract

Decrease of nicotine, other tobacco alkaloids (OTAs), and carcinogenic tobacco specific nitrosamines (TSNAs) is important for health. To reduce these compounds, we selected MRE and G-box elements of promoters of four *Nicotiana tabacum Jasmonate ZIM-domain* (*NtJAZ)* repressor genes as targets to screen transcription factors (TFs) and then identified Arabidopsis Production of Anthocyanin Pigment 1 (PAP1) and Transparent Testa 8 (TT8) for a novel regulation design. Electrophoretic mobility shift, cross-linked chromatin immunoprecipitation, and dual luciferase analyses demonstrated that PAP1, TT8, PAP1/TT8, and the PAP1/TT8/Transparent Testa Glabra 1 (TTG1) complex bound to and activated *NtJAZ* promoters. Stacked *PAP1* (P) and *TT8* (T), which were introduced to commercial tobacco cultivars, Narrow Leaf Madole (NL) and KY171 (KY), upregulated four *NtJAZs*, downregulated seven/eight nicotine biosynthetic genes, and reduced nicotine and nornicotine in transgenic red P+T-NL and P+L-KY plants. Field trials of red P+T-NL, P+L-KY, and PAP1 tobacco genotypes and metabolic quantification showed significant reduction of nicotine and four OTAs in most or all cured leaves. Three TSNAs and the fourth one were diminished by 63-92% and 30-74% in contents in all leaves of P+T-NL and P+L-KY genotypes, respectively. The total content of each TSNA was significantly reduced by 25-60% in leaves of PAP1 plants. Taken together, this regulation designed crossing from Arabidopsis anthocyanin to tobacco alkaloid pathway is defined as a distant pathway-cross regulation (DPCR). The proof of

concept of DPCR unearths new functions of two known TFs, new activators of *NtJAZs*, and a negative regulation pathway of nicotine biosynthesis in red tobacco.

Key words: Arabidopsis, alkaloids, anthocyanin, flavonoid, bHLH, JAZ, MYB, PAP1, TT8, nicotine, tobacco, and transcription factor

Introduction

Nicotine is the main alkaloid of tobacco (*Nicotiana tabacum*) and is harmfully addictive (Arany et al., 2018; Grando, 2014; Greillier et al., 2018; Lee et al., 2017; Sanner and Grimsrud, 2015; Santoro et al., 2019; Thorndike and Rigotti, 2009; Tidey et al., 2018). Other harmful tobacco alkaloids include nornicotine, anabasine, anatabine, and myosmine (Figure 1 A-B). Carcinogenic tobacco specific nitrosamines (TSNAs) are derived from these alkaloids. Five common TSNAs in tobacco smoke are N-nitrosonornicotine (NNN), nicotine-derived nitrosamine ketone (NNK), N'-nitrosoanatabine (NAT), N-nitrosoanabasine (NAB), and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) (Hecht, 1998; Hecht et al., 2016; Konstantinou et al., 2018). Given that NNN is a severe carcinogen, the Food and Drug Administration has proposed a rule that the average level of NNN must be less than 1.0 ppm (dry weight) in all final smokeless tobacco products (FDA, 2017). A global long-term effort is to eliminate these metabolites in tobacco products.

To date, numerous achievements have been accomplished in elucidating the biosynthesis of nicotine and other tobacco alkaloids (Dawson, 1945; Dewey and Xie, 2013; Ladesic and Tso, 1964; Mizusaki et al., 1973; Stepka and Dewey, 1961; Tso and Jeffrey, 1956, 1957). Nicotine has pyridine and pyrrolidine rings that are biosynthesized from two distinct pathways (Figure 1 A). Three genes, *AO*, *QS*, and *QPT2*, encode three enzymes that catalyze steps from aspartate to nicotinic acid (pyridine) through quinolinate and nicotinic mononucleotide intermediates (Figure 1 A) (Dewey and Xie, 2013; Shoji and Hashimoto, 2013). QPT1, a QPT2 paralog, is not involved in the formation of pyridine (Shoji and Hashimoto, 2011a). Five genes, *ODC*, *ADC*, AIH, *PMT*, and *MPO*, have been proved to encode enzymes that catalyze steps from ornithine or arginine to 4-methylaminobutanal, which spontaneously forms N-methyl-Δ^1-pyrrolinium cation with a pyrrolidine ring (Figure 1 A) (Dewey and Xie, 2013; Shoji and Hashimoto, 2013). Although the condensation mechanism of the two rings remains open for further studies, two genes, namely *A622* and *BBL*, have been shown to involve these last steps (DeBoer et al., 2009; Dewey and Xie, 2013; Hibi et al., 1994; Kajikawa et al., 2009; Kajikawa et al., 2011; Lewis et al., 2015; Shoji and Hashimoto, 2013). The demethylation of nicotine catalyzed by a CYP82E4 and its homologs forms nornicotine (Chakrabarti et al., 2008; Gavilano and Siminszky, 2007; Lewis et al., 2010). The mechanisms of the anatabine and anabasine formation are unknown (Dewey and Xie, 2013). One hypothesis is that these two alkaloids also

result from the BBL catalysis of nicotinic acid. The other is that anabasine is derived from lysine, a different pathway (Figure 1 A) (Dewey and Xie, 2013). Four main TSNAs are NNK, NNN, NAT, and NAB (Figure 1 A-B), which result from the nitrosation of nicotine, nornicotine, anatabine, and anabasine during the leaf curing process, respectively. To date, the mechanisms related to the nitrosation still require further investigation (Dewey and Xie, 2013). Furthermore, the regulation mechanisms of the nicotine pathway are intensively elucidated. Myelocytomatosis oncogene (MYC) and APETALA 2/ethylene response factor (AP2/ERF) are two transcription factors (TFs) that directly bind to promoters of pathway genes to regulate the nicotine biosynthesis (De Boer et al., 2011; Dewey and Xie, 2013; Shoji and Hashimoto, 2011b, 2013; Zhang et al., 2012). It is also elucidated that jasmonate (JA), a plant hormone, essentially regulates the nicotine biosynthesis (Baldwin et al., 1997; Imanishi et al., 1998; Shoji et al., 2000; Xu and Timko, 2004). In the past decade, the regulation mechanism of the nicotine biosynthesis by JA was elucidated by the demonstration of a co-receptor complex consisting of coronatine insensitive 1 (COI1) (Xie et al., 1998), a Skp/Cullin/F-box (SCF^{COI1}) complex, and JA ZIM-domain (JAZ) repressor (De Boer et al., 2011; Shoji and Hashimoto, 2011b; Shoji et al., 2008; Zhang et al., 2012). When roots lack the bioactive isoleucine-JA (JA-Ile), two *N. tabacum* JAZ repressors, NtJAZ1 and NtJAZ3, bind to NtMYC2 to turn off the nicotine biosynthesis. In addition, a recent study revealed additional 13 NtJAZs, in which NtJAZ7 and 10 were shown to associate with the regulation of the nicotine

biosynthesis (Zhang et al., 2019). When JA-Ile exists, it induces the interaction of SCF^{COI1} and JAZ to form a complex, which leads to the ubiquitination of JAZs to release NtMYC2. The freed NtMYC2 then activates the biosynthetic pathway of nicotine (Figure 1 A) (Shoji and Hashimoto, 2011b; Shoji et al., 2008; Zhang et al., 2012). Moreover, NtMYC2 is a positive feedback regulator that activates the transcription of *NtJAZ1* and *NtJAZ3 (Shoji and Hashimoto, 2011b; Shoji et al., 2008)*. This feedback loop regulation is considered being the only mechanism that controls the transcription of JAZs and JA-signaling pathway in all plants (Chini et al., 2007; Goossens et al., 2017; Howe et al., 2018). To date, based on those pathway and regulatory genes, on the one hand, antisense and RNAi have been reported to down-regulate gene expression (such as *BBL*, *ODC*, *QPT*, and nicotine demethylase gene) to reduce nicotine, nornicotine, or TSNAs in leaves (DeBoer et al., 2013; Lewis et al., 2015; Xie et al., 2004b; Zhao et al., 2016). On the other hand, these multiple successes also revealed a challenge that the reduction of nicotine was traded with the increase of anatabine and other alkaloids (Chintapakorn and Hamill, 2003; DeBoer et al., 2011). These successful data apparently indicate that efforts are necessary to diminish TSNAs, nicotine, and other alkaloids simultaneously.

The biosynthesis of anthocyanin has been comprehensively elucidated from both model and non-model plants. MYB and bHLH TFs together with a WD40 protein form MBW complexes to regulate the biosynthesis of anthocyanins (Saigo et al.,

2020; Shi and Xie, 2014; Springob et al., 2003). *Arabidopsis Production of Anthocyanin Pigment 1* (*PAP1*) encodes a R2R3-MYB TF, namely MYB75 (Borevitz et al., 2000). *Arabidopsis Transparent Testa 8* (*TT8*) encodes a bHLH TF, namely bHLH42 (Nesi et al., 2000). *Arabidopsis Transparent Testa Glabra 1* (*TTG1*) encodes a WD40 protein (Walker et al., 1999). The functions of these three genes in Arabidopsis and their homologs in other plants have been intensively characterized in genetics and biochemistry (Xu et al., 2014). PAP1 and TT8 bind to MRE and G-box elements in promoters of anthocyanin pathway genes, respectively, such as *DFR*, *ANS* and *3-GT* (Thevenin et al., 2012; Xu et al., 2013; Zhu et al., 2015). The current dogma is that on the one hand, PAP1, TT8, and TTG1 form a master regulatory MYB-bHLH-WD40 (MBW) complex to activate anthocyanin biosynthesis in plants (Figure 1 C) (Baudry et al., 2004; Ramsay and Glover, 2005; Shi and Xie, 2014; Shi and Xie, 2011); on the other hand, Arabidopsis JAZs interact with PAP1, TT8 and the MBW complex to control anthocyanin biosynthesis in the presence of jasmonate (JA) and other phytohormones (Qi et al., 2011; Qi et al., 2014). Moreover, *PAP1* has been used to create red crop varieties for value-added traits , such as tobacco (He et al., 2017; Xie et al., 2006), tomato (Butelli et al., 2008; Zuluaga et al., 2008), hop (Gatica-Arias et al., 2012), and canola (Li et al., 2010). All these studies reported that PAP1 alone could activate anthocyanin biosynthesis in different plants. In contrast with all knowns, whether PAP1, TT8, and their homologs have other regulatory functions for new values remain for investigations.

This study was a proof of concept investigation to use known Arabidopsis transcription factors and tobacco *JAZ* repressor genes to innovate new technologies to reduce nicotine, OTAs, and TSNAs simultaneously. Given that NtJAZ1 and NtJAZ3 were demonstrated to be two main repressors as described above, we hypothesized that they were appropriate targets to screen TFs for a high expresion to downregulate nicotine biosynthesis. Then, we analyzed cis-regulatory elements of all *NtJAZ* promoters and identified four *NtJAZ* genes with MRE(s) and/or G-box(es). Meanwhile, we analyzed Arabidopsis TFs that have a regulatory fucntion to bind to these two types of elements. Then, PAP1 and TT8, two master regulators of Arabidopsis anthocyanin biosynthesis, were identifed for the design of a distant pathway-cross regulation (DPCR) from the anthocyanin to nicotine pathway. The constructed DPCR significantly diminished harmful nicotine, four OTAs, and four TSNAs in tobacco plants, simultaneously. Moreover, the resulting data disclose new regulatory functions of both PAP1 and TT8.

Results

MRE and G-box elements identified in the promoters of *NtJAZ1, 3, 7,* and *10*

NtJAZ1 and NtJAZ3 are two repressors of nicotine biosynthesis (Shoji and Hashimoto, 2011b; Shoji et al., 2008; Zhang et al., 2012). In recent, 13 *NtJAZs* were mined from the tobacco genomes, showing that *NtJAZ7*, and *NtJAZ10* were also associated with the regulation of nicotine biosynthesis (Zhang et al., 2019). We identified the promoter sequences of 13 *NtJAZs* from the genome of NC90 tobacco curated at NCBI and named these as *NtJAZ1pro* through *NtJAZ13pro*. One or 1.5 kb nucleotides in the proximal upstream from the start codon (ATG) were obtained for sequence characterization. Sequence analysis revealed that *NtJAZ1pro*, *NtJAZ3pro*, and *NtJAZ7pro* had both MRE and G-box elements, while *NtJAZ10pro* only had G-box elements (Figures S1-S2). Next, we cloned the promoter sequences of *NtJAZ1pro* and *NtJAZ3pro* from two commercial dark tobacco cultivars, Narrow Leaf Madole (NL) and KY171 (KY). Their sequences were identical to those of *NtJAZ1pro* and *NtJAZ3pro* from TN90 (Figure S1), therefore, we used the same names in this report. Sequence analysis further substantiated the MRE and G-box elements in *NtJAZ1pro* and *NtJAZ3pro*. *NtJAZ1pro* has two MRE and two G-box elements (Figure 2 A and Figure S1). The two MRE elements are MRE1: ACCCACC at position -180 and MRE2: ACCCCAC at position -172. The two G-boxes are CACGTG at positions -232 and AACGTG at -395. *NtJAZ3pro* has one MRE3 element: AACTACC at position -169 and one G-box: CACGTG at position -960 (Figure 2 A and Figure S1).

EMSA and ChIP-Q-PCR show the binding of PAP1 and TT8 to the *NtJAZ1* and *NtJAZ3* promoters

EMSA assay was carried out to test whether PAP1 and TT8 could bind to the MRE and G-box elements of *NtJAZ1*$_{pro}$ *and NtJAZ3*$_{pro}$ *in vitro*. The R2R3-MYB binding domain (amino acids 1-178) of Arabidopsis PAP1 was cloned to the pDEST-His-MBP vector, in which it was fused with the His-MBP-Tag at the N-terminus (Figure S3 A). The bHLH binding domain (amino acid 359-443) of TT8 was cloned to the PRSF-Dute vector, in which it was fused with the His-Tag at the N-terminus (Figure S3 B). Two constructs were introduced to *E. coli* to induce recombinant proteins, which were further purified with a His-tag purification system (Figure S3 A-B). Biotin probes were prepared for MRE1 (ACCCACC), MRE2 (ACCCCAC), MRE3 (AACTACC), and G-box (CACGTG). Both competitive and non-competitive probes were used as controls for EMSA assay. The resulting data showed that the recombinant R2R3 binding domain directly bound to three MRE element probes and the recombinant bHLH binding domain directly bound to the G-box (Figure 2 B). The binding activity of both recombinant R2R3 and bHLH binding domains was competed by competitive probes but not competed by non-competitive probes (Figure 2 B). These data demonstrate that both PAP1 and TT8 bind to the *NtJAZ1*$_{pro}$ *and NtJAZ3*$_{pro}$.

ChIP-Q-PCR analysis was performed to examine whether PAP1 and TT8 could bind to *NtJAZ1pro* and *NtJAZ3pro in vivo*. In ChIP experiments, *PAP1* was fused at the 5-end of *GFP* in the binary vector pGWB5 to obtain pGWB5-PAP1-GFP, which was introduced into leaves of wild type NL and KY171 seedlings for transient expression via injection of *Agrobacterium*. *TT8* was fused at the 5-end of an *HA-tag* sequence in the binary vector PK2GW7 to obtain PK2GW7-TT8-HA-tag, which was also introduced into leaves of wild type NL and KY171 seedlings for transient expression by injection of *Agrobacterium*. The IP experiments of PAP1 and TT8 were performed by using GFP and HA monoclonal antibodies. The ChIP PCR results showed that the amplified *NtJAZ1* and *NtJAZ3* promoter fragments was increased 5.0-40.4 times by PAP1 and 6.1.8-37.5 times by TT8 (Figure 2 C).

Dual-luciferase assay demonstrates the activation of *NtJAZ1* and *NtJAZ3* promoters by PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1

Dual-luciferase assay was completed to demonstrate whether PAP1 alone, TT8 alone, PAP1 and TT8 together, and PAP1/TT8/TTG1 (WD40) together could activate the *NtJAZ1* and *NtJAZ3* promoters in tobacco. PAP1, TT8, and TTG1 were used as effectors. The ORF sequences of *PAP1*, *TT8*, *TTG1*, and *GFP* were respectively cloned into the PCB2004 plasmid to generate four constructs, in which each gene was driven by a 35S-promoter (Figure 2 D). The *NtJAZ1pro* and *NtJAZ3pro* were cloned

into the plasmid pGreenII-0800 to drive the luciferase gene to obtain *NtJAZ1pro*-luciferase and *NtJAZ3pro*-luciferase reporters (Figure 2 D). Effectors and reporters were incubated for dual-luciferase assays. Promoters alone and a binary vector were used as controls. The resulting data showed that compared with controls, PAP1 and TT8 alone enhanced the activity of *NtJAZ1pro* 3.8 and 7.2 times, respectively, and increased the activity of *NtJAZ3pro* 1.9 (slightly) and 2.1 (significantly) times, respectively (Figure 2 E). PAP1 and TT8 together enhanced the activity of *NtJAZ1pro* and *NtJAZ3pro* approximately 4.0 and 6.3 times (Figure 2 E). Furthermore, PAP1, TT8, and TTG1 together increased 12.7 and 15.6 times of the activity of both *NtJAZ1pro* and *NtJAZ3pro*, much higher than PAP1 alone, TT8 alone, and PAP1 and TT8 together (Figure 2 E). These results demonstrate that PAP1 and TT8 alone, PAP1 and TT8 together, PAP1-TT8-TTG1 complex transcriptionally activate both *NtJAZ1* and *NtJAZ3*.

PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1 do not activate promoters of *ODC2* and *PMT* of tobacco

NtODC2 and *NtPMT2* are two key upstream pathway genes of the nicotine biosynthesis. We obtained 1.0 kb long sequences for these two gene's promoters from the tobacco genome database curated at NCBI. The sequences of the *NtODC2* promoter does not have MRE and G-box elements, while the promoter of *NtPMT2* has

an MRE-like element (AACAACC at position -530 and a G-box like (CACGTT at position -186) element (Figure S4).

EMSA experiments were performed to test whether PAP1 and TT8 could bind to MRE-like and G-box-like elements. The EMSA results indicated that the TT8 binding domain could bind to the G-box-like (CACGTT) element. However, when a competitive probe was added, the binding signal was lost (Figure S5 A). Furthermore, compared with the positive control G-box elements from *NtJAZ1* and *NtJAZ3*, the resulting data showed that the signal from G-box like element was much weaker (Figure S5 A). This result indicates that TT8 can weakly bind to the promoter of *NtPMT2*. When the MRE-like element probe was prepared for EMSA, the resulting data showed that PAP1 could not bind to the MRE-like probe (AACAACC), while PAP1 strongly bound to the positive control MRE1 and MRE3 probes of *NtJAZ1* and *NtJAZ3* promoters (Figure S5 B).

Dual-luciferase assay and ChIP-Q-PCR were performed to test whether PAP1 and TT8 could activate the *ODC2* and *PMT2* promoters. The dual-luciferase assay data did not show that PAP1 alone, TT8 alone, PAP1/TT8 together, and the PAP1/TT8/TTG1 complex could activate the activity of the two promoters (Figure S6 A). Meanwhile, the ChIP-Q-PCR data showed that PAP1 and TT8 could not enrich binding elements of the two promoters *in vivo* (Figure S6 B). These data demonstrate

that PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1 cannot activate the promoters of these two key pathway genes.

Red genotypic P+T-NL and P+T-KY tobacco plants and characterization of anthocyanin and proanthocyanidin biosynthesis

More than 20 T0 transgenic plants were generated from each of both wild type Narrow Leaf Madole (NL) and KY171 (KY) varieties with a stacked *PAP1* (P) and *TT8* (T) binary vector (Figure S7 A, B, C, E, and G-H). In addition, vector control transgenic plants were generated (Figure S7 D and F). The first 12 red T0 transgenic plants from the two cultivars were tagged as P+T-NL1 to 12 and P+T-KY1 to 12, respectively. The first 12 T0 vector control transgenic plants from the two cultivars were tagged as NL1 to 12 and KY1 to 12, respectively. These T0 plants were selected to obtain seeds to screen T1 and homozygous T2 progeny (Figure S7 I-M), during which the copy number of the transgenes was determined. Five P+T-NL lines (1-5) and five P+T-KY lines (1-5), which highly expressed *PAP1* and *TT8* (Figure S8), were demonstrated to contain one single copy of the stacked two transgenes. Furthermore, qRT-PCR was performed to understand the expression of *CHS*, *CHI, F3H, F3'H*, *DFR*, and *ANS* involved in the biosynthesis of both anthocyanin and proanthocyanidin (PA) and *3-GT* involved in the glycosylation of anthocyanidins (Figure 1 C) in T2 seedlings (Figure S9). The resulting data showed that the

expression levels of these seven genes were up-regulated 2-1000 times (Figure S10 A-B) in pink roots and red leaves of the P+T-NL and P+T-KY genotypes. Further UV-spectrometry-based measurements showed that the roots and leaves of analyzed P+T-NL1 and P+T-KY1 genotypes produced anthocyanins, while those of control plants did not (Figure S11). It was noticeable that the levels of anthocyanins were significantly higher in leaves than in roots of both P+T-NL1 and P+T-KY1 plants (Figure S11). To understand this differentiation, the expression of *ANR* specifically involved in the colorless PA formation (Xie et al., 2003) was analyzed. The resulting data showed that the expression of *ANR* was significantly activated in roots but not in leaves of the two red genotypes (Figure S10). To substantiate this, root extracts of both P+T-NL1 and P+T-KY1 seedlings were treated with DMACA. A bluish color was immediately developed from the reaction (Figure S12 A-B), showing the presence of PAs. However, this bluish color was not developed from the extracts of both P+T-NL1 and P+T-KY1 leaves and wild type roots and leaves, indicating the lack of PAs. Further measurement of DMACA-treated samples at 640 nm showed that the ABS values were 5-6 times higher in the root extracts than in the leaf extracts of P+T-NL1 and P+T-KY1 and in the leaf and root extracts of wild type (Figure S12 C-D). These data showed that the less red pigmentation in roots resulted from the conversion of red anthocyanidins to colorless PAs by ANR. In addition, the less red pigmentation in roots might result from the lack of light and secretion of anthocyanins.

Upregulation of *NtJAZ1, 3, 7,* 10, and downregulation of the nicotine biosynthesis in red P + T-NL and P + T-KY plants

Both transcriptional and metabolic profiling were performed to characterize the biosynthesis of nicotine in homozygous T2 seedlings (Figure S9) of different genotypes selected, including P+T-NL1, P+T-NL2, P+T-NL3, P+T-KY1, P+T- KY 2, and P+T-KY3. Real time qRT-PCR was performed to characterize the transcriptional profiles of 16 genes (Figure 1 a) involved in the nicotine biosynthesis. The expression levels of *NtJAZ1, NtJAZ3, NtJAZ7, and NtJAZ10* were significantly increased in roots of both P+T-NL and P+T-KY genotypes (Figure 3 A-B and K-L, Figure S13). The expression levels of *PMT1*, *PMT2*, *MPO*, *QPT2*, *A622*, *BBL* and *ERF189* were significantly reduced in roots of both P+T-NL (Figure 3 D-J) and P+T-KY plants (Figure 3 N-T). The expression levels of *OCD2* was not altered in the P+T-NL1, 2, and 3 lines (Figure 3 C) but significantly decreased in the P+T-KY1, 2 and 3 lines (Figure 3 M). The expression levels of *MYC1a*, *MYC1b*, *MYC2a*, and *MYC2b* were not altered in roots of the P+T-NL and P+T-KY plants (Figure S14). HPLC-MS analysis showed that the contents of nicotine, nornicotine, and the total amount of the two compounds in roots and leaves of analyzed P+K-NL1, P+K-NL2, P+K-NL3, P+K-KY1, P+K-KY2, and P+K-KY3 plants were significantly reduced 1.5-3.4, 1.3-2.5, and 1.9-3.3 fold, respectively (Figure 4).

Reduction of nicotine, nornicotine, anatabine, anabasine, myosmine, and four TSNAs in air-cured leaves of red P+T-NL1 and P+T-KY1 tobacco plants

Field trials that follow the commercial production practices of tobacco are essential to examine the effects of new varieties or genes on the contents of nicotine, other tobacco alkaloids, and TSNAs. We planted T2 homozygous P+T-NL1 and P+T-KY1 progeny and their corresponding wild-type tobacco variety in the field at the research station in Oxford, North Carolina (Figure 5 A-C and Figures S15-S18). The field practice followed the protocols of both industry tobacco and GMO tobacco growth. Tobacco leaves were air-cured in a contained ventilation barn (Figure 5 D, Figures S18-S19).

Quantification with GC-FID showed significant reduction of nicotine, nornicotine, anabasine, anatabine, and myosmine in most or all cured leaves of red P+T-NL1 and P+T-KY1 tobacco plants compared to those of non-transgenic controls. The contents of nicotine were significantly reduced by 45-51% in all cured leaf groups of P+T-NL1 (Figure 6 A) and by 19-30% in upper three groups of cured leaves of P+T-KY1 (Figure 6 B). The contents of nornicotine were significantly decreased by 39-44% and 30-40% in all cured leaf groups of T-P-NL1 and P+T-KY1, respectively (Figure 6 C

and D). The contents of anabasine, anatabine, and myosmine were significantly reduced by 27-45%, 55-66%, and 19-25% in all cured leaf groups of P+T-NL1 (Figure 6 E, G, and I), respectively, and were significantly reduced by 27-45%, 32-56%, and 29-35% in all cured leaf groups of P+T-KY1, respectively (Figure 6 F, H, and J).

Quantification with HPLC-QQQ-MS/MS showed that the contents of four carcinogenic TSNAs were significantly decreased in most or all leaf groups of P+T-NL1 and P+T-KY1 plants. The content of NNK was reduced by 38-72% and 30-74% in four groups of cured P+T-NL1 and P+T-KY1 leaves, respectively, in which the reduction in three leaf groups of each genotype was significant (Figure 7 C and D). The content of NNN was significantly decreased by 63-74% in all four groups of cured P+T-NL1 leaves and by 67-79% in three groups of cured P+T-KY1 leaves (Figure 7 A and B). The content of NNN in the B3 leaf group of P+T-KY1 was reduced by 63% with a slight significance. The content of NAT was significantly decreased by 77-92% and 70-80% in all leaf groups of P+T-NL1 and P+T-KY1, respectively (Figure 7 E and f). The content of NAB was significantly decreased by 72-80% and 60-77% in all leaf groups of P+T-NL1 and P+T-KY1, respectively (Figure 7 G and H). More fundamentally, the contents of NNN in all leaf groups were diminished to less than 0.5 ppm, which was two-fold lower than the FDA's proposed 1.0 ppm in any smokeless tobacco products (FDA, 2017; Konstantinou et al., 2018).

Furthermore, the contents of NNN, NNK, NAT, and NAB were summed to obtain the total TSNA contents in each leaf group and entire plant leaves. The contents of total TSNAs in each group of wild type NL and KY171 leaves were 1.6-2.2 ppm and 2.2-2.6 ppm, while these values were significantly reduced to be 0.4-0.6 ppm and 0.6-1.0 ppm in each group of P+T-NL1 and P+T-KY1 leaves (Figure 7 I-J). The contents of total TSNAs in the entire plant leaves of wild type NL and KY171 were 7.07 ppm and 9.01 ppm, while these values were reduced to be 2.67 ppm and 3.47 ppm in those of P+T-NL1 and P+T-KY1 (Figure 7 K-L). Furthermore, the total contents of tobacco alkaloids in each group of leaves were summed from nicotine, four other alkaloids, and total TSNAs. The resulting data showed that the contents of total alkaloids were significantly decreased by 44-51% in four leaf groups of P+T-NL1 and 22-35% in the upper three group leaves of P+T-KY1 (Figure S20 A-B).

Reduction of nicotine, nornicotine, anatabine, anabasine, myosmine, and four TSNAs in flue-cured leaves of the red PAP1 tobacco

To understand whether PAP1 alone could reduce nicotine, other alkaloids, and TSNAs, we investigated our red PAP1 tobacco genotype (Figure 1 C), a homozygous red variety generated from Xanthi by *PAP1* (Xie and Dixon, 2005). Gene expression analysis showed that the expression of *PAP1* upregulated two *TT8* homologs, *NtAn1a*

and *NtAn1b* (Bai et al., 2011) (Figure S21). This result indicated that *PAP1* alone regulated tobacco metabolism via activating endogenous *N. tabacum TT8* (*NtTT8*) homologs.

We conducted field trials in 2011 and 2012. The field design (Figure S22 A-B) and cropping management were the same as described above. When leaves were harvested, they were harvested into five groups (Figure S22 C-D) and then flue-cured in barns (Figure S22 E). Quantification with HPLC-ESI-MS showed the reduction of nicotine, nornicotine, anabasine, and anatabine. The levels of nicotine were significantly or slightly reduced by 10-50% and 5-25% in five group leaves in 2011 and in 2012, respectively (Figure S23 A-B). The levels of nornicotine, anabasine, and anatabine were significantly or slightly reduced in all or most group leaves by 25-70% and 5-55%, 30-60% and 5-20%, and 5-60% and 10-40% in 2011 and 2012, respectively (Figure S23 A-B). In addition, the total level of each alkaloid was summed from all leaf groups of the entire PAP1 tobacco plant. The resulting data showed that the total levels of nicotine, nornicotine, anabasine, and anatabine in the entire plant were significantly reduced by 30%, 45%, 50%, and 40% in 2011 and 10%, 20%, 10%, and 30% in 2012 respectively (Figure S24 A-B).

More importantly, the contents of NNN, NNK, NAB, and NAT were reduced in cured leaf groups of red PAP1 tobacco in 2011 and 2012 (Figure S25 A-B). The contents of NAB were reduced by 20-70% and 15-50% in all leaf groups in two years. The contents of NAT were reduced by 50-65-% and 35-70% in all leaf groups in two years. The contents of NNK were reduced by 5-50% in all leaf groups in 2011 and 5-60% in four of five leaf groups in 2012. The contents of NNN were reduced by 20-60% and 25-60% in all leaf groups in two years. The average content of NNN was diminished to level less than 0.5 ppm in most leaf groups (Figure S25 A-B), two-fold lower than the FDA's proposed 1.0 ppm in any smokeless products. The contents of NNK, NAT, and NAB were also reduced to less than 1.0 ppm in all leaf groups (Figure S25 A-B). The total contents of each NNN, NNK, NAB, and NAT were summed for all leaf groups of each plant. The resulting data showed that the total contents of each TSNA were significantly reduced by 25-60% (P-value < 0.05) in 2011 and 2012 (Figure S25 A-B). The total contents of four TSNAs were summed for all leaf groups of each plant. The resulting data showed that the total contents of four TSNAs were significantly reduced by 57% and 43% (P-value < 0.05) in 2011 and 2012, respectively (Figure S26).

Discussion

Given that the pathway of nicotine does not exist in Arabidopsis, two anthocyanin TFs are distant from the tobacco alkaloid pathway. Herein, we constructively term the designed regulation to be a *Distant Pathway-Cross Regulation* (DPCR) of PAP1 and TT8. This DPCR definition characterizes that PAP1, TT8, and their complex formed with the tobacco WD40 homolog negatively regulate the tobacco alkaloid biosynthesis via activating *NtJAZ* repressor genes. Accordingly, this definition distinguishes from the master regulatory function of PAP1, TT8, and their MBW complex that activates the anthocyanin biosynthesis in Arabidopsis (Baudry et al., 2006; Borevitz et al., 2000; Gonzalez et al., 2008; Zhu et al., 2015) or red tobacco shown here via directly activating most or late pathway genes, such as DFR, ANS, and GTs (Ramsay and Glover, 2005; Shi and Xie, 2014; Shi and Xie, 2011; Zhu et al., 2015) (Figure S10). Based on all data, we summarize a model to characterize the DPCR and the regulation of tobacco alkaloid biosynthesis in red tobacco plants (Figure 8). In this model, PAP1 and TT8 anchor to the tobacco NtTTG1 (WD40) to form a heterogeneous MBW complex, which has a dual regulatory function that regulates tobacco metabolism. First, the MBW complex directly activates the anthocyanin biosynthesis (function I), which is intensively elucidated. Second, the MBW complex upregulates four *NtJAZ1*, *NtJAZ3*, *NtJAZ7*, and *NtJAZ10*, which increase NtJAZ repressors (function II). Next, the increased NtJAZs bind to more MYC2 and block the release of this TF. Then, the reduction of MYC2 results in downregulating *EFR* and six or seven nicotine pathway genes (Figure 3). Finally, the

biosynthesis of nicotine, OTAs and four TSNAs is significantly diminished in red tobacco plants (Figures 4, 6, and 7). In addition, this model shows the fine-tuned regulation of the nicotine biosynthesis formed by JA, JAZ and MYC2. To date, PAP1 or PAP1/TT8 and their homologs have been introduced to a few crops to engineer anthocyanins for value-added traits, such as tomato (Butelli et al., 2008; Zuluaga et al., 2008), tobacco (Bai et al., 2011; He et al., 2017; Xie et al., 2006), hop (Gatica-Arias et al., 2012), canola (Li et al., 2010), and rose (Zvi et al., 2012). To our knowledge, whether and how PAP1 and TT8 can regulate non-anthocyanin or non-flavonoid pathways in other plants remains investigated. It was interesting that PAP1 transgenic rose plants were reported to produce more than 6.5-time scent terpenoids (Zvi et al., 2012). Based on our findings, it is valuable to investigate whether the increase of scent terpenoids result from a similar DPCR in rose. Moreover, different TF families are a large part of plant genomes (Feng et al., 2020; Millard et al., 2019; Pireyre and Burow, 2015; Skinner et al., 2010), our findings suggest that it is necessary to investigate DPCR of other TFs to fully understand their regulatory functions. In particular, investigations of potential DPCR of TFs likely create value-added plant traits for agriculture and human health, as shown in our data that carcinogenic TSNAs were significantly reduced in red tobacco plants.

The DPCR formed by PAP1/TT8 TFs and *NtJAZs* shows a promising application in enhancing efforts for tobacco harm reduction. To date, it has been a challenging task

to reduce harmful nicotine, OTAs, and TSNAs, simultaneously. In the past decades, multiple previous studies had been undertaken to develop technologies to reduce these compounds, such as gene suppression or silencing technologies with anti-sense or RNAi of pathway genes (Figure 1) including *PMT* (Chintapakorn and Hamill, 2003), *ODC* (DeBoer et al., 2011), *QPT* (Xie et al., 2004a), *NDM* (Lewis et al., 2008), A622 (DeBoer et al., 2009), *BBL*(Lewis et al., 2015), and others (Dewey and Xie, 2013). Actually, most studies by manipulating these genes reported the decrease of nicotine content in tobacco plants grown in the greenhouse or growth chambers. Moreover, the suppression of *BBL* and *QPT* practiced in the field obtained promising reductions of nicotine. However, these previously reported technologies showed that continuous efforts are necessary to decrease nicotine, TSNAs, and OTAs simultaneously due to unwanted metabolic tradeoffs or limited suppression on tobacco alkaloid formation. One problem was that the decrease of nicotine led to either the increase of anatabine or non-reduction of other tobacco alkaloids (Chintapakorn and Hamill, 2003; DeBoer et al., 2011). The other problem was that the reduction of harmful compounds was limited. For instance, an RNAi of A622 could reduce NNN and total TSNAs, but could not decrease nicotine, anabasine, anatabine, and other TSNAs (Lewis et al., 2008). A gene silencing of *BBL* was reported to effectively reduce the nicotine content four-fold (Lewis et al., 2015), however, this technology was not shown to reduce TSNAs and other alkaloids. Moreover, different *MYC2* isomers were selected to manipulate tobacco alkaloids. However, the resulting data showed that the

reduction of nicotine and other alkaloids to a practical level remained for more studies (Wang et al., 2015; Zhang et al., 2012). This difficulty likely results from the feedback regulation of the nicotine biosynthesis. As documented in the literature, the regulation of the nicotine biosynthesis is characterized with three cohort layers, the presence/absence of an active JA form, the JA and JAZ signaling pathway, and the release of MYC2 for activation of ERF189 and pathway genes (Figure 1 A). In particular, this regulation is fine-tuned by the interaction of MYC2 and JAZs proteins and the MYC2 feedback activation of *JAZs* (Shoji and Hashimoto, 2011b, 2013; Zhang et al., 2012). When the ubiquitination of JAZ repressors occurs to release MYC2, the freed MYC2 feeds back to activate the transcription of *NtJAZs*. It is apparent that this feedback loop forms a seamless regulation mechanism to control the biosynthesis of nicotine and other tobacco alkaloids in roots. To innovate new technologies for tobacco harm reduction, we started with analyzing promoter sequences of most known genes involved in the biosynthesis of nicotine. We identified that the promoter sequences of *NtJAZ1*, *NtJAZ3*, *NtJAZ7*, and *NtJAZ10* contained G-box and/or the MRE(s) (Figure 2 A and Figures S1 and S2), the binding cis-elements of TT8 and PAP1, respectively. Based on these, we succeeded in using these two TFs to create the DPCR to downregulate the biosynthesis of tobacco alkaloids. The contents of nicotine and nornicotine were significantly reduced in non-topped seedlings grown in the greenhouse (Figure 4). More significantly, three years of field trials showed that the contents of nicotine, four OTAs, and four TSNAs were

simultaneously reduced in most or all cured leaf groups and entire plant leaves of P-T-NL, P-T-KY, and PAP1 genotypes (Figures 6 and 7, and Figures S20 and S23-26). The reduction of the NNN content to a level less than 0.5 ppm in most of red tobacco leaves (Figure 7 A-B, Figure S25 A-B) is particularly significant to human health, because it is a severe tobacco carcinogen and its content in any smokeless tobacco products is limited to 1.00 pm proposed by FDA. In addition, other TSNAs were reduced to a level lower than 0.5 ppm in most or all leaves of all three genotypic red tobacco plant. Given that we mainly focused on characterizing tobacco alkaloids, we have not analyzed other trait changes in our three years' field trial. Although we have not observed obvious vegetative growth alternation because of plant topping, it is valuable to study other traits in the future. Taken together, this PAP1-TT8-JAZ based DPCR allows the reduction of nicotine, four TSNAs, and four OTAs, simultaneously. We anticipate that this proof of concept provides a new platform to eliminate tobacco carcinogens to a non-harmful level in the future.

Experimental Procedures

Tobacco varieties

Six genotypic tobacco (*Nicotiana tabacum*) cultivars were used in this study. Two commercial dark tobacco varieties were Narrow Leaf Madole (NL) and KY171 for

smokeless products (Andersen et al., 1990 ; Pearce et al., 2015). Two red transgenic genotypes were generated from NL and KY171 described below. Two others were *N. tabacum* Xanthi (an oriental tobacco variety) and a red PAP1 tobacco variety that is a novel homozygous red Xanthi variety (Figure 1 C) created from the overexpression of the *Arabidopsis PAP1* (MYB75) (Xie et al., 2006). *Nicotiana benthamiana* was grown for dual-luciferase assay.

Growth conditions in the phytotron

Seedlings of all genotypes were grown in the phytotron for isolation of DNA and RNA, cloning, sequencing, gene expression profiling, and metabolite analysis. Plants were grown both in water and in soil in 10 by 10 cm (height x diameter) size pots. The growth conditions consisted of a constant temperature of 28±3ºC and a 16/8-h (light/dark) photoperiod with a light intensity of 200 μmol m^{-2} s^{-1}. For hydroponic growth, after seed germination in soil, five-day seedlings were transferred to foam float cubes, which were placed in trays filled with Hoagland nutrient solution. Plants then grew 45 days for harvesting roots and leaves, which were immediately frozen in liquid nitrogen and then stored in a -80ºC freezer until use for isolation of DNA and RNA and other experiments. Meanwhile, seedlings grown in pot soil were grown 45 days to collect leaf tissues. *Nicotiana benthamiana* plants were grown in soil in the same phytotron conditions for dual-luciferase experiments. Leaves of 35 days old plants were used as material for dual-luciferase assays described below.

Cloning of *NtJAZ1, NTJAZ3, PMT2,* and *ODC2* promoters from Narrow Leaf Madole and KY171 cultivars

The promoter sequences of tobacco *NtJAZ1, NtJAZ3, NtJAZ7, NtJAZ10, PMT2*, and *ODC2* were identified from the genomic sequences of the NT90 cultivar curated at NCBI (https://www.ncbi.nlm.nih.gov/). The sequences identified were 1000 bp or 1500 bp. Based on these sequences, primers (Table S1) were designed for PCR to amplify *NtJAZ1, NtJAZ3, PMT2*, and *ODC2* promoters from the genomic DNA of wild type NL and KY171 tobacco plants, and then the amplified genomic DNA fragments were cloned to pEASY-T1 for sequencing. Briefly, the DNA fragments from PCR were purified using a gel-extraction Kit (Thermo Fisher, Waltham, MA) following the manufacturer's instruction and then ligated to the pEASY-T1 plasmid using a T4-ligase system (10 µl reaction system: 1 µl T4 buffer, 1 µl T4-ligase, 1.0 µl pEASY-T1 linear plasmid and 7.0 µl purified PCR product). The resulting recombinant plasmid was transformed to the competent *E. coli* Top10 cells, which were screened on agar-solidified LB medium with 50 mg/l ampicillin. Positive colonies were selected for sequencing at Eton Bio (Durham, NC, USA). All obtained promoter sequences were 1000 or 1500 bp (Figures S1, S2, and S4). Analysis of promoter sequences and identification of regulatory elements such as AC- and AC-like elements (Figures S1, S2, and S4) were completed with the PLACE (http://www.dna.affrc.go.jp/PLACE/) and PlantPan (http://plantpan.itps.ncku.edu.tw/)

tools. The *NtJAZ1*, *NtJAZ3*, *PMT2*, and *ODC2* promoters were used for electrophoretic mobility shift assay, dual-luciferase, and ChIP assays described below.

Electrophoretic mobility shift assay of PAP1 and TT8 binding to promoters

To perform electrophoretic mobility shift assay (EMSA), we designed gene-specific primers (Supporting Table S1) for PCR to amplify the binding domains of both *PAP1* and *TT8*. The amplified *PAP1* binding domain from the N-terminus was 534 bp, which included the *R2R3 MYB* domain sequences for Gate-way cloning. The amplified *TT8* binding domain franking the *bHLH* region was 252 bp, in which both BamHI and SaclI restriction sites were added to two ends for further cloning. After the confirmation of sequence accuracy, three expression vector systems, pRSF-Dute, pet-Dute, and pDest-His-MBP (Li et al., 2017), were comparatively tested to induce soluble recombinant proteins. We selected the pDest-His-MBP vector, a Gateway cloning system, for the *R2R3* domain of *PAP1*. This vector franks an MBP-tag and a His-tag (5- His-MBP-3) in the N-terminus. We introduced it into the entry vector pDonr221 through BP reaction (Gateway™ BP Clonase™ Enzyme Mix, Carlsbad, CA) and then performed LR exchange reaction (Gateway™ LR Clonase™ Enzyme Mix, Carlsbad, CA) to clone the *R2R3* binding domain sequence to the 5-end of His-tag-MBP-tag in the pDest-His-MBP vector by following the manufacturer's protocol. The resulting PAP1-N-pDest-His-MBP vector with ampicillin selection was

transformed to the competent cells of *E. coli* strain BL21 (DE3). Positive colonies were screened for protein induction described below.

We found that the pRSF-Dute vector with a His-tag in the N-terminus was appropriate for expressing the bHLH domain of TT8 after several vectors were tested. In order to clone the *bHLH* domain to the 5-end of His-tag in the pRSF-Dute vector, the 252 bp PCR product of the *bHLH* domain described above and the pRSF-dute PRSF vector were digested with BamHI and Sac I, then were ligated using NEB T4 ligase by following the manufacturer's protocol (NEB, MA, USA). The resulting expression vector, TT8-bHLH-pRSF-dute with kanamycin selection, was transformed to the competent cells of *E. coli* strain BL21 (DE3). Positive colonies were screened for protein induction.

One positive colony of *E. coli* (BL21/DE3) containing PAP1-N-pDest-His-MBP or TT8-bHLH-pRSF-dute was inoculated to 200 ml liquid LB medium contained in a 500 ml E-flask. The flasks were placed on a shaker and then incubated for 10 hrs at a speed of 250 rpm/min at 37°C. When the concentration of *E. coli* reached to an OD value of 0.6 measured at 600 nm, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added into the suspension culture up to 500 μM. The *E. coli* culture was continuously incubated overnight at 28°C and then harvested into 50 ml tubes by 10 min of centrifugation at 4000 rpm and 4°C. The remained *E. coli* pellet was thoroughly

suspended in autoclaved double deionized water, followed by centrifugation as described above. This wash step was repeated once. The washed pellet was thoroughly suspended in 100 mM Tris-HCl buffer (pH 7.5) and then lysed with ultrasonication with a 400 W power for 20 min (every 3 sec with a 3 sec interval) on ice. The lysed mixture was centrifuged at 4°C for 30 min. The resulting supernatant containing soluble proteins was transferred to a new 50 ml tube for purification. Protein purification was completed with a His-tag Protein Purification kit (Clonetech, Catalog 635659, Mountain View, CA) by following the manufacturer's protocol. Finally, the recombinant PAP1-His-MBP-Tag and TT8-bHLH-His-Tag proteins were collected in pH7.0 Tris-HCl buffer. The resulting purified proteins (Supporting Figure S3) were used for EMSA assay described below.

Probes were prepared for EMSA of PAP1 and TT8. Based on the regulatory elements obtained from the four promoter sequences and PAP1 binding features, four sequences were selected to prepare different probes for PAP1, MRE1: ACCCACC, MRE2: ACCCCAC, MRE3: AACTACC, and MRE-like: AACAACC. Based on the four promoter sequences and TT8 binding features, two sequences probes, G-box: CACGTG and G-BOX-like: CACGTT, were prepared for TT8. These six probes were synthesized with a biotin labeling using a Pierce™ Biotin 3' End DNA Labeling Kit (Thermo Fisher, Waltham, MA).

The DNA-binding reactions between probes and two TFs were carried out using a LightShift™ Chemiluminescent EMSA Kit (Thermo Fisher, Waltham, MA) according to the manufacturer' protocol. The reactions were carried out in 20 µl of mixture, which was composed of 2.0 µg purified protein and 0.5 mM biotin-labeled DNA probe reaction buffer, 50 mM KCl, 5% (vol/vol) glycerol, and 1.0 µg poly(deoxyinosinicdeoxycytidylic) acid (polydI:dC). After the binding reactions were completed in 20 min at room temperature, all reaction products were loaded on 6% (g/100 ml) nondenaturing polyacrylamide gels, which were placed into a gel box for electrophoresis at 160 V for 60 min. Proteins were then transferred to Hybond-N+ nylon membranes (Thermo Fisher, Waltham, MA). After five times of strict washing, the band shifts were detected using Chemiluminescent Nucleic Acid Detection Module system (Thermo Fisher, Waltham, MA) by following the manufacturer's manual. Binding images were photographed using a pre-cold CCD camera (SYNGENE GBOX, MD, USA).

Dual luciferase assays

Four reporter and four effector vectors were developed to perform dual luciferase assays that were completed to analyze the regulatory activity of PAP1 alone, TT8 alone, PAP1 and TT8 together, and PAP1/TT8/TTG1 (WD40) together in activating *NtJAZ1*, *NtJAZ3*, *PMT2*, and *ODC2* promoters. Steps are described in details below.

First, the pGreenII-0800 vector was used to develop four reporter vectors. This vector contains multiple restriction enzyme sites immediately at the N-terminus of a firefly-luciferase gene, such as BamHI, NotI, and EcoR V. This vector also contains a *NPTII* gene for kanamycin-based screening. To clone promoters to the pGreenII-0800 vector, primers (Supporting Table S1) were designed to add two restriction enzyme sites at the two ends to amplify each promoter sequence. The primer pairs designed for *NtJAZ1, NtJAZ3, ODC2, and PMT2* promoters included BamHI and NotI, EcoRV and BamHI, and EcoR V and Not I restriction enzymes sites, respectively. Each promoter sequence was amplified by PCR with the same thermal programs (Supporting Table S1) and then purified as described above. The pGreenII-0800 vector and promoters of *NtJAZ1*, *NtJAZ3*, *PMT2*, and *ODC2* were digested by specific restriction enzymes as designed. The digested pGreenII-0800 and each promoter sequence were ligated using T4 ligase. The resulting ligation products were transformed to competent *E. coli* Top10 cells. Positive *E. coli* colonies were obtained from screening on agar-solidified LB medium supplemented with 50 mg/ml kanamycin. One positive colony from each ligation was inoculated in 50 ml liquid LB medium supplemented with 50 mg/l kanamycin contained in 250 ml an E-flask, which was placed on a shaker at a speed of 250 rpm for an overnight culture at 37ºC. After harvesting suspension cultures, plasmids were purified with a gel purification kit. Finally, four reporter vectors, namely $NtJAZ1_{pro}$-pGreenII-0800, $NtJAZ3_{pro}$-pGreenII-

0800 (Figure 2 D), *ODC2*$_{pro}$-pGreenII-0800, and *PMT2*$_{pro}$-pGreenII-0800 were established for introduction to *Agrobacterium* described below.

Next, four effector vectors were developed from the PK2GW7 vector (He et al., 2017), to which the ORFs of *GFP*, *PAP1*, *TT8*, and *TTG1* were cloned. Each gene was driven by a 35S-promoter. First, the forward and reverse primers designed for amplification of the ORFs of these genes were added an attB1 (forward primer) and an attB2 (reverse primer) adapter (Supporting Table S1). PCR was performed to amplify each ORF with an attB1 at the 5-end and an attB2 adapter at the 3-end. Each PCR product was purified using a gel-extraction kit (Thermo Fisher, Waltham, MA). The purified promoters were ligated to the pDonr207 plasmid with BP reactions. These ligations generated plasmids pDonr207-GFP, pDonr207-PAP1, pDonr207-TT8, and pDonr207-TTG1, which were then introduced to competent *E. coli* Top10 cells to identify positive colonies screened on LB medium containing 50 mg/ml gentamicin. The *GFP*, *PAP1*, *TT8*, and *TTG1* ORFs in the pDonr207 plasmids were then cloned to the expression vector PK2GW7 via LR reactions. The resulting recombinant PK2GW7-GFP, PK2GW7-PAP1, PK2GW7-TT8, and PK2GW7-TTG1 plasmids were introduced to competent *E. coli* Top10 cells to select positive colonies on LB medium containing 50 mg/l spectinomycin, from which positive colonies harboring each recombinant plasmid identified. One positive colony for each recombinant plasmid was then inoculated in 50 ml liquid LB medium supplemented with 50 mg/l

spectinomycin contained in 250 ml an E-flask, which was placed on shaker for overnight culture as described above. The *E. coli* cultures were harvested to isolate each recombinant effector plasmid.

Then, four reporter and four effector vectors were introduced to *Agrobacterium tumefaciens* strain GV3101, respectively. For each reporter vector, an enhancer vector pSoup+P19 was used as a cooperator for vector pairing. Accordingly, four pairs, pSoup-P19 and the $JAZ1_{pro}$-pGreenII-0800 reporter, pSoup-P19 and $JAZ3_{pro}$-pGreenII-0800, pSoup-P19 and $PMT2_{pro}$-pGreenII-0800, and pSoup-P19 and $ODC2_{pro}$-pGreenII-0800, respectively, were transformed to competent *Agrobacterium* GV3101 by an electroporation shock method with 2000V voltage. The transformed *Agrobacterium* GV3101 cells were screened on LB medium supplemented with 50 mg/l rifampicin and 50 mg/l kanamycin at 30°C. The positive colonies harboring each reporter vector were named as reporter Agrobacterium GV3101. Using the same method, four effector vectors, *GFP*-PK2GW7, *PAP1*-PK2GW7, *TT8*-PK2GW7, and *TTG1*-PK2GW7, were transformed to competent *Agrobacterium* GV3101, respectively. The transformed *Agrobacterium* GV3101 cells were screened on LB medium supplemented with 50 mg/l rifampicin and 50 mg/l spectinomycin at 30°C. The positive colonies harboring each effector vector was named as effector Agrobacterium GV3101. Each type of reporter and effecter Agrobacterium GV3101 were inoculated into 50 ml liquid LB medium in 250 ml E-flasks, respectively, and

cultured for 48 hrs on a shaker with 200 rpm at 28°C under dark. Then, 100 µl suspension culture of each reporter Agrobacterium GV301 and effecter Agrobacterium GV301 was inoculated to 50 ml fresh liquid LB medium supplemented with antibiotics contained in a 250 ml E-flask. Each flask was placed on the same shaker for continuous culture. When OD values reached to 0.6 measured at 600 nm, cultured effector or reporter GV301 cells were harvested to 50 ml tube and centrifuged at 4000 rpm for 10 min. The resulting *Agrobacterium* pellets were suspended in infiltration buffer (10 mM pH5.5 MES, 10mM $MgCl_2$, 150 µM acetosyringone) to an OD value of 0.2-0.3. The re-suspended *Agrobacterium* cells were then incubated at room temperature for two hrs prior to transient expression experiments described below.

Last, the activated effector and reporter *Agrobacterium* GV3101 cells were used to inoculate leaves of *N. benthamiana* for transient expression assay. Prior to inoculation, 100 µl suspension cells of one reporter *Agrobacterium* GV3101 were completely mixed with 450 µl suspension cells of one corresponding effecter *Agrobacterium* GV3101. The resulting mixture formed an effector-reporter vector pair, such as PAP1 vs. $NtJAZ1_{pro}$. Meanwhile, four young leaves on one 35-day old *N. benthamiana* plant were used as four replicates for inoculation. On each leaf, 210 µl mixed suspension cells was infiltrated at three locations, each with 70 µl. Five independent plants as biological replicates were infiltrated with each effector (TF) and

reporter (promoter) pair (e.g. PAP1 vs. *NtJAZ1*$_{pro}$). After infiltration, all plants were grown in the phytotron for four days prior to the analysis of dual-luciferase detection using a Dual-Luciferase Reporter Assay System kit (Promega, Madison, USA). Briefly, 1.0 cm diameter-size leaf discs at inoculated locations were punched into a 1.5 ml tube and then homogenized in 500 µl Passive Lysis Buffer (included in the kit). The homogenized mixture was centrifuged for 10 min at 12,000 rpm at 40°C. The resulting supernatant was transferred to a new 1.5 ml tube and then diluted 50 times with the Passive Lysis Buffer. Ten µl diluted extract was mixed with 40 µl Luciferase Assay Buffer. The firefly luminescence (LUC) from samples was immediately recorded 5.0 sec on a GloMax 20/20 luminometer (Promega, Madison, USA) followed by 10 sec intervals. Then, the reaction was stopped by adding 40 µl Stop and Glow Buffer, and immediately recorded for one second luminescence using renilla (REN) luminescence. The ratios of LUC to REN were calculated. For each pair of a reporter (promoter) and an effector (TF), five biological replicates were performed to obtain LUC to REN ratio values. Five biological replicates were also performed for the empty pGrenn-II-0800 and *GFP* gene as reporter and effecter controls.

Cross-linked chromatin immunoprecipitation (ChIP) assays

A gateway cloning was used to develop two plasmids, in which GFP was fused to the

C-terminus of PAP1 and TT8 for genetic transformation. A pair of primers (Supporting Table S1) was designed to amplify the *PAP1* ORF without its stop codon. The resulting PCR product was purified as described above and then cloned into the pDonr221 vector via a BP reaction. The *PAP1* fragment was further cloned to the upstream of *GFP* in the destination vector pGWB5 via a LR reaction. This cloning step generated a recombinant vector pGWB5-PAP1-GFP, in which *PAP1-GFP* was driven by a 35S promoter. The pGWB5-PAP1-GFP vector was introduced to the competent *E. coli* Top10 cells. Positive colonies were selected on LB medium containing 50 mg/l kanamycin. One colony was used for suspension culture on the same shaker as described above. Cultured *E. coli* cells were harvested to isolate the pGWB5-PAP1-GFP binary plasmid, which was further introduced to competent *Agrobacterium* GV301 cells. To develop a binary vector expressing *TT8*, a pair of primers (Supporting Table S1) was designed for PCR. The reverse primer was designed to include a HA-tag sequence. The resulting PCR product was cloned into the plasmid PK2GW7 as described above. Next, the resulting binary vector PK2GW7-TT8-HA-tag was introduced to competent *Agrobacterium* GV301 cells. *Agrobacterium* GV301 cells harboring *PAP-GFP* or *TT8-HA-tag* cDNA were activated and then injected to leaves of 5-week old wild type NL and KY171 seedlings that were grown in the phytotron. After injection, plants were allowed to grow 72 hrs. The infected leaves were then harvested to liquid nitrogen to isolate nuclei.

Nuclei were isolated from infected and un-infected (control) leaves. Frozen leaf tissue (500 mg) was ground into fine powder in liquid nitrogen and then suspended thoroughly in 25 ml nuclei isolation buffer (HEPES pH7.6, 1M sucrose, 5 mM KCl, 5mM MgCl, 5 mM EDTA, 1% formaldehyde, 14 mM 2ME, 0.6 % triton X-100, 0.4 M PMSF). The mixture was kept at room temperature for 10 min of cross-linking, then added 1.7 ml 2.0 M glycine to stop the cross-linking. The cross-linked mixture was filtered through a gauze to remove residues. The flow-through containing nuclei was centrifuged at 3000 × g for 10 min. After the removal of the supernatant, the remaining nuclei were suspended in isolation buffer (HEPES pH7.6, 1M sucrose, 5 mM KCl, 5 mM MgCl, and 5 mM EDTA) and mixed appropriately. Then, 800 µl 15% Percoll solution (15% Percoll, 10 mM HEPES, 1M sucrose, 5 mM KCl, 5 mM MgCl, 5 mM EDTA) was added to the mixture and gently vortexed, followed by centrifugation at 3000 x g for 5 min. After the removal of supernatant, the nuclei pellet was suspended in nuclei lysis buffer (50 mM Tris-HCl pH7.5, 1% SDS, 10 mM EDTA) for 30 min to release chromatin. The chromatin suspension was then sonicated for 3 min (every 4 sec with 9 sec interval) on ice in a Sonic Dismembrator (Fisherbrand™ Model 505 Sonic Dismembrator, Thermo Fisher, Waltham, MA). This sonication sheared DNA (chromatin)-protein into 250-1000 bp fragments, which were examined by gel electrophoresis. Then, the chromatin mixture was centrifuged at 13,000 × g for 3.0 min to obtain the supernatant for IP.

GFP monoclonal antibody GF28R and HA-Tag monoclonal antibody (2-2.14) (ThermoFisher, USA) were used for ChIP. First, the IP assay mixture was composed of 100 µl supernatant (chromatin mixture), 900 µl ChIP dilution buffer (1 % triton X-100, 1 mM EDTA, 15 mM Tris-HCl pH7.5, 150 mM NaCl), and 5 µl HA-antibody or GFP-antibody contained in a 1.5 ml tube. The tube was incubated for 4 hrs at 4 ºC. Next, the cross-linked DNAs were immunoprecipitated by the HA-antibody or GFP-antibody. The resulting immune complexes were washed three times with 0.5 ml lysis buffer (0.05 M HEPES, 250 mM NaCl, 1 mM EDTA, 1 % triton X-100, 0.1 % sodium deoxycholate (m/v), 0.1%SDS, 10 mM sodium butyrate, 0.001 mM PMSF, 1x cocktail), followed by three times of washing with 0.5 ml LNDET buffer (250 mM LiCl, 1% NP40, 1 % sodium deoxycholate (m/v), 1 mM EDTA), and three times of washing with 0.5 ml TE buffer (10 mM Tris-HCl pH7.5, 1 mM EDTA). These washes removed free antibodies and non-specific chromatins. Then, the DNA-protein complex was suspended in 0.5 ml TE buffer in a 1.5 ml tube. The same volume of phenol/chloroform (1:1, v/v) was added into the tube and appropriately mixed. The tube was centrifuged at 12,000 ×g for 10 min. The upper aqueous phase was transferred to a new 1.5 ml tube and then added 0.1 volume 3.0 M sodium acetate. The tube was gently vortexed and centrifuged at 12,000 × g for 10 min. After the removal of the supernatant, the precipitated DNA was air-dried and dissolved in 30 µl autoclaved double deionized-water. Last, the purified DNA was used as templates for

ChIP-Q-PCR analysis. In addition, a positive input DNA (DNA fragments without IP reaction) was used as a positive control. Primers used for amplifying four promoters and thermal programs were as described above (Supporting Table S1).

Synthesis of T-DNA, development of binary constructs, and genetic transformation of two dark tobacco varieties

PAP1 and *TT8* of *A. thaliana* encode a R2R3-MYB and bHLH TF, respectively. A T-DNA cassette was designed and then synthesized to stack *PAP1* and *TT8* for their coupled overexpression. From 5'- to 3' end, the cassette was composed of *attL1*, the *PAP1* cDNA (NP_176057), *NOS* (terminator), *35S* promoter, the *TT8* cDNA (CAC14865), and *attL2* (Figure S7 A). The synthesized sequence was introduced to the entry vector pUC57 and then cloned into the destination binary vector pK2GW7 by attL × attR combination (LR) reaction with LR clonase. The resulting binary vector, namely PAP1-TT8-pK2GW7 (Figure S7 B), was further introduced into the competent cells of *Agrobacterium tumefaciens* strain GV3101. A positive colony was selected for genetic transformation of tobacco. In addition, the pK2GW7 vector was used for genetic transformation as control. These constructs were transformed into two globally commercial dark tobacco varieties, NL and KY171, respectively. Detailed steps (Figure S7 C) for transformation and selection of T0 transgenic plants followed our protocols reported previously (He et al., 2017). Transgenic plants from

NL and KY171 were labeled as P+T-NL and P+T-KY, respectively. More than 20 T0 transgenic plants were obtained for each variety and grown in the greenhouse to select T0 seeds, 12 of which were further selected on medium supplemented with antibiotics to screen antibiotic resistant T1 progeny. A 3:1 ratio of seed germination to seed death indicated one single copy of T-DNA that was in the genomic DNA of T0 plants. This ratio was observed in the T1 progeny of T0 plants. In this study, we focused on lines with one single copy of transgenes. A large number of seeds were obtained from the T1 progeny of each line. Seeds from these types of T1 progeny were screened on an agar-solidified MS medium supplemented with 50 mg/l kanamycin. If all seeds from a T1 plant could germinate T2 seedlings, it indicated that this T1 plant was a homozygous line. Accordingly, red homozygous T2 plants (Figure S7 M) were obtained and used for growth in the greenhouse and field trial. In addition, PCR and qRT-PCR were performed to demonstrate the presence of transgene in the genome of T0 and T2 transgenic plants and the expression of both *PAP1* and *TT8* in both P+T-NL and P+T-KY plants.

DNA and RNA extractions

Steps are listed in Method S1.

Polymerase chain reaction and quantitative reverse transcript-polymerase chain reaction

Steps are listed in Method S2.

Assays of anthocyanin and proanthocyanidin in seedlings

Methods were followed our anthocyanin (He et al., 2017) and proanthocyanidin assay protocols (Wang et al., 2020). Steps are listed in Method S3.

Field trial of P+T-NL1 and P+T-KY1 T2 progeny and air curing

Homozygous T2 progeny of P+T-NL1 and P+T-KY1 were planted at the research station in Oxford, North Carolina in 2015. In addition, NL and KY171 plants were planted as controls. Prior to planting transgenic plants in the field, we applied for GMO permits from APHIS at USDA for this field trial. North Carolina State University is licensed to grow transgenic tobacco plants in the field. Every single step followed the protocol requested in the license. Field planting followed the protocol of commercial tobacco production in North Carolina, including selection and designing of field, planting, phenotypic observation, field management, topping, harvest of leaves, and air curing (Supporting Figure S15).

The field used to grow transgenic plants was isolated from other fields by a forest. Three field plots were designed as three planting replicates (Supporting Figure S16 A). Red tobacco plants and control plants were grown in rows side by side. Two or three rows of commercial K326 tobacco plants were grown beside red and control plants as a buffer zone in order to enclose transgenic plants (Supporting Figure S16 A).

Field trails included three steps: seed germination and seedling growth in foam float trays (Supporting S16 B and Table S2) in a greenhouse (containment), transportation of seedlings, and field planting. Seeds were planted in foam float trays in hydroponic conditions (Supporting Figure S16 B). During seed germination, three plots were selected at the research station, supplemented with 560 kg. ha^{-1} of 8-8-24 base fertilizer, and then treated with clomazone and carfentrazone-ethyl/sulfentrazone for weed control. After two months of growth, seedlings were transported to and planted in the field (Supporting Figure S17). Fifty plants for each genotype were planted in one row in each field plot. In addition, wild type NL and KY171 plants were planted side by side as a control (Supporting Figure S17). One plot was planted 200 plants in four rows including 50 plants for each of P+T-NL1, P+T-KY1, NL, and KY171 genotypes. Three plots were planted as field replicates. Accordingly, each genotype was planted 150 plants in total. Four rows of wild type K326 tobacco plants were planted between two neighbor plots and around plots as the buffer zones. The

management of plant growth in the field followed a protocol developed for commercial plants in the research station. Three plots were side dressed with 168 kg ha^{-1} of 15.5-0-0 (Nitrogen-phosphate-potassium) fertilizer. In the first week of July, when flower buds started to develop from shoot apex, all plants were topped and kept with 16 leaves. After topping, plants were treated with maleic hydrazide to control suckers.

In the first week of August, topped plants were cut from the base and hung up upside down on a truck, and then transported to a ventilation barn for air curing at the research station (Supporting Figure S18 A-B). Leaves from the base to the top of plants were labelled as 1-16. After the harvest, plant residues were completely cleaned up from the field (Supporting Figure S18 C-E) and transported to the laboratory for autoclaving. The air curing took one to two months to dry all leaves (Supporting Figure S19 A-B). Air-cured leaves of each plant were harvested beginning from the base to the top positions. The first four base leaves (#1-4) from 10 plants were collected together as one biological sample. This group of leaves was labelled as B1 (Supporting Figure S19 C-D). Therefore, B1 had five biological replicates in one plot. In three plots, B1 had 15 biological samples. Likewise, the second four leaves (# 5-8) were collected and labelled as B2. The third four leaves (#9-12) were labelled as B3. The top four leaves (#13-16) were labeled as B4. After the harvest, all leaves were further dried at 70°C in an oven for 24 hr. The dried leaves were ground into fine

powder in a stainless-steel blender (Conair, East Windsor, NJ) and were used for alkaloid content analysis as described below.

Field trial of red PAP1 tobacco plants and flue-curing

PAP1 tobacco is the progeny of a homologous red PAP1 Xanthi plant that is generated by the overexpression of *PAP1* (Xie et al., 2006). Field design, seed germination, plantation, field management, and plant growth management were the same as described above. The field trials for PAP1 tobacco were conducted for two years. In 2011 and 2012, seedlings were planted in the field on May 11 and May 1, respectively. The different planting dates were selected because of weather conditions. In addition to red PAP1 tobacco plants, wild type Xanthi was planted as control, and two commercial cultivars, K326, and NC7, were planted as protective buffer zones. Three plots were designed as three replicates and each plot was planted 100 plants (Supporting Figure S22 A-B). In total 300 red PAP1 tobacco plants were grown in the field. Plants were topped to keep with 15 leaves on July 15, 2011 and July 10, 2012, when flower buds started to develop from shoot apexes. Leaves from the base to the top were numerated #1- #15, which were arranged into five groups, B1: leaves #1-3, B2: leaves 4-6, B3: #7-10, B4: #11-12, and B5: #13-15 (Supporting Figure S22 C). Xanthi tobacco is a Turkish cultivar, thus, its leaf harvest protocol is

different from that used for dark tobacco described above. In addition, the dates of leaf harvest were different in 2011 and 2012 due to weather conditions. In 2011, leaves were harvested from plants from September 12 to 28. On September 12, 2011, the group B1 leaves were harvested from 1-50 plants. Leaves from five plants were pooled together as one biological replicate. Accordingly, each plot had 10 biological replicates for the B1 group. In total, 30 biological samples were obtained for the B1 group from three plots. Likewise, 30 biological samples were obtained for B2, B3, B4, and B5 groups. After excised from plants, each leaf was immediately cut longitudinally into two halves from the median vein (Supporting Figure S22 D). One half with the median vein was flue-cured in a barn at the research station. The other half was immediately frozen in liquid nitrogen and transported to the laboratory and stored in a -80°C freezer. For each plot, ten biological replicates were prepared for both flue-cured and non-cured leaves for late experiments. Likewise, five groups of uncut entire leaves were harvested from 51-100 plants. Leaves from five plants were pooled together as one biological sample. These leaves were used as a control of cut leaves. The group B2, B3, B4, and B5 leaves were harvested on Sept. 18, 14, and 28. Steps were the same as those for the group B1 leaves. In 2012, the group B1, B2, B3, B4, and B5 leaves were collected on Aug. 9, Aug. 16, Aug. 24, Sept. 2, and Sept. 13. Steps of leaf harvest were the same as those in 2011. After one week of flue-curing, dried leaves were stored in a dry room and then ground into fine powder for nicotine,

nornicotine, and TSNA analysis as described below. Any plant residue was completely cleaned up from the field.

Determination of TSNAs in cured leaves of field-grown P+T-NL1 and P+T-KY1 tobacco plants with HPLC-QQQ-MS/MS

Accurate quantification of tobacco specific nitrosamines **(**TSNA) was completed with high performance liquid chromatography coupled with triple quadrupole tandem mass spectrometer (HPLC-QQQ-MS/MS). This analysis was performed using industrial protocols on an Agilent 1200 series HPLC System coupled with AB Sciex Triple quadrupole mass spectrometer (API 5000 with Electrospray) in RJ Reynolds. Four TSNAs, N-nitrosonornicotine (NNN), N-nitrosoanatabine (NAT), N-nitrosoanabasine (NAB), and nicotine-derived nitrosamine ketone (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, NNK) in air-cured leaves, were quantified. One hundred milligrams of dry leaf powder were extracted in an aqueous ammonium acetate solution (100 mM aqueous ammonium acetate containing four deuterium analogs for TSNA) and filtered through disposable PVDF syringe filters into autosampler vials. The extract was separated in a Phenomenex Gemini C18 column with 3.0 µm particle 2.0 x150 mm. The injection volume was 2.0-5.0 µl. TSNAs were detected by multiple reaction monitoring (MRM) of the precursor ion to a product ion specific for each compound. Quantification was achieved using an internal standard calibration

comprised of ten points. A separate internal standard was used for each analyte by using a mixture of four stable isotope-labeled analytes. Results are reported in units of ng/g (ppb). Data were determined to be acceptable if the correlation coefficient for the calibration curve was greater than 0.99, standards were within 85% to 115% of expected concentration values, check solution values were within 75% to 125% of target, peak shape and resolution were acceptable (based on historical data), and values for the QC samples were within established control limits. The instrument parameters for MS, HPLC gradient, the quadrupole mass spectrometer parameters, reagents and standards are listed in in Supporting Tables S3-S6.

Determination of nicotine and other tobacco alkaloids in cured leaves of field-grown P+T-NL1 and P+T-KY1 tobacco plants by GC

Analysis of gas chromatograph (GC) equipped with a flame ionization detector (FID) was performed using a standard industry method to quantify nicotine and other tobacco alkaloids in RJ Reynolds. In this study, analysis of nicotine and other tobacco alkaloids was performed on an Agilent 6890 gas chromatograph (GC) equipped with a flame ionization detector (FID) and an Agilent 7683 automatic sampler. One hundred milligrams of air-cured dry leaf powder were alkalinized in 2 mM sodium hydroxide (NaOH). This solution was extracted with methyl-tert-butyl ether (MTBE) spiked with an internal standard using a wrist-action shaker. The mixture was allowed to

separate, and the resulting MTBE layer was transferred to an autosampler vial for analysis of nicotine, anabasine, nornicotine, and myosmine with GC-FID. Quantification was achieved using an internal standard calibration comprised of six points. Data were determined to be acceptable if the correlation coefficient of the calibration curve was greater than 0.998, the response factors (RF) were consistent, the percent relative standard deviation (%RSD) for RF was equal or less than 5%, the check solution values were within 10% of target, values for the quality control (QC) samples were within established control limits, and chromatograms had appropriate identification of peaks. The instrument parameters, the oven program, and reagents and conditions used in alkaloid analysis are listed in Supporting Table S7-S9. The results were reported as percentiles (%, g/g, dry weight).

Determination of TSNAs and tobacco alkaloids in flue-cured leaves of field-grown red PAP1 tobacco plants by high performance liquid chromatography-electrospray ionization-mass spectrometry (HPLC-ESI-MS)

HPLC-ESI-MS analysis was performed on a 2010EV LC/UV/ESI/MS instrument (Shimadzu, Japan). One gram of powdered sample was extracted in 20 ml of methanol in a flask placed on a shaker for one hour at room temperature. The samples were centrifuged twice at 4000 rpm and aliquots of the supernatant were transferred to a 2.0 ml glass vial. The samples were separated on an Eclipse XDB-C18 analytical

column (250 mm ×4.6 mm, 5 µm, Agilent, Santa Clara, CA, USA). The mobile phase solvents were composed of 1% acetic acid in water (solvent A; HPLC-grade acetic acid and LC–MS grade water) and 100% LC–MS grade acetonitrile (solvent B). To separate metabolites, the following gradient solvent system, with ratios of solvent A to B, was used: 90:10 (0–5 min), 90:10–88:12 (5–10 min), 88:12–80:20 (10–20 min), 80:20–75:25 (20–30 min), 75:25–65:35 (30–35 min), 65:35–60:40 (35–40 min), 60:40–50:50 (40–55 min), and 50:50–10:90 (55–60 min). The column was then washed for 10 min with 10% solvent B. The flow rate and the injection volume were 0.4 ml/min and 20 µl, respectively. The total ion chromatograms of positive electrospray ionization were recorded from 0 to 60 min by mass spectrum detector and mass spectra were scanned and stored from m/z of 120–1,600 at a speed of 1,000 amu/s. Standards of nicotine, nornicotine, anabasine, NAB, NAT, NNAL, NNK, and NNN (Sigma, St-Louis, MO) were used to establish standard curve with a coefficient value at least 98%.

Determination of nicotine and nornicotine in leaves of seedlings with HPLC-ESI-MS

Seedling of P+T-NL1, P+T-KY1, vector control, and wild type plants were hydroponically grown in the phytotron. Leaves and roots of 45-day old seedlings (Supporting Figure S9) were used to analyze nicotine and nornicotine in both leaves

and roots. Fresh tissues were ground into fine powder in liquid nitrogen and dried in a 45°C oven. One hundred mg of powdered sample was extracted with 2.0 ml methanol. Other steps were as described above. The analysis of nicotine and nornicotine was performed with HPLC-TOF-MS on Agilent 6520 time-of-flight LC-MS/MS instrument (Agilent Technologies, Santa Clara, CA, USA). Mobile solvents and column used for separation of metabolites were solvent A and B described above. The gradient solvent system was composed of ratios of solvent A to B: 95:5 (0–5 min), 95:5 to 90:10 (5–10 min), 90:10 to 85:15 (10–15 min), 85:15 to 45:55 (15–20 min), 45:55 to 25:75 (20–25 min), 25:75 to 95:5 (25–26 min). After the last gradient step, the column was equilibrated and washed for 10 min with solvents A: B (95:5). The flow rate was 0.4 mL/min. The injection volume of samples was 5.0 µL. The dry gas flow and the nebulizer pressure were set at 12 L/min and at 50 psi, respectively. Metabolites were ionized with the negative mode. The mass spectra were scanned from 100 to 3000 m/z. The acquisition rate was three spectra per second. Other MS conditions included fragmentor: 150 V, skimmer: 65 V, OCT 1 RF Vpp: 750 V, and collision energy: 30 psi. Standard curves were developed for both nicotine and nornicotine. The results were reported as mg/g, dry weight.

Statistical analysis

Values for each group sample such as B1 were averaged from samples of three plots. One-way ANOVA was used to evaluate the statistical significance of the contents of nicotine, nornicotine, anabasine, NNN, NNK, NAT, and NAB between red and wild type leaves. Standard deviation was used to show the variation of metabolite contents. For gene expression, protein analysis, nicotine and nornicotine contents in seedlings, all values are averaged from four or five biological replicates. Student t-test was performed to evaluate statistical significance. P-values less than 0.05, 0.01, and 0.001 mean significant, very significant, and extremely significant.

Acknowledgement

R.J. Reynolds Tobacco Company provided funding (#571426) and analytical support for metabolite analysis of field samples in this project. We are grateful to Dr. Jannell Rowe for analyzing nicotine, TSNAs and other alkaloids in P+T-NL and P+T-KY samples. We are thankful to our colleague professor Ramsey Lewis for providing us seeds of KY171 and Narrow Leaf Madole cultivars from the tobacco germplasm bank at NCSU and for suggestions about field trial regulation of transgenic tobacco plants and leaf curing. We are grateful to staff at Research Station in Oxford, North Carolina for helping planting, sucker control, irrigation, and field fertilization as well as providing both flue-curing barns and air-curing facility. We are thankful to Drs. Emmett Hiatt and Dr. Jannell Rowe from R. J. Reynolds for critically reading this

manuscript. We are grateful to Professors Ian Mackay and Zhao-Bang Zeng for advising statistical analysis.

Competing financial interests

The authors declare no competing financial interests.

Authors' contribution

DYX perceived and designed the entire project, developed all experimental plans, selected field, supervised analysis of all data and figure preparation, drafted and finalized the manuscript. ML designed and performed EMSA, dual-luciferase, CHIP, gene expression profiling, promotor cloning, PCR, qRT-PCR, and anthocyanin experiments, analyzed all data and prepared figures, and participated manuscript drafting. XH performed gene synthesis, construct development, genetic transformation and selection of P+T-transgenic plants, screening of homozygous T2 progeny, waterbed nursery of P+T seedlings in greenhouse, field preparation, planting, management of plant growth, plant harvest, air-curing, and sample preparation for HPLC-MS analysis, sample preparation for HPLC-QQQ-MS/MS and GC-FID experiments to analyze nicotine, nornicotine, and TSNAs in RJ Reynolds. and result drafting. CLH completed waterbed nursery of PAP1 seedlings, field preparation, planting, management of plant growth, plant harvest, flued curing, HPLC-ESI-MS, data analysis, and result drafting. YZ analyzed nicotine and nornicotine in seedlings. YD extracted RNA and performed qRT-PCR. SL, HX, YL, YJ, and MM were involved in experiments including selection of T2 progeny, greenhouse experiments, planting, plant harvest, and collection of field data, sampling, and sample preparation for metabolite analysis.

Supporting information

Supporting materials and growth conditions, Method S1-S12, 27 supporting figures and 9 supporting tables.

27 Supporting Figure legends

Figure S1 Promoter sequences of *NtJAZ1* and *NtJAZ3* and identification of cis-elements

Figure S2 Promoter sequences of *NtJAZ7* and *NtJAZ10* and identification of cis-elements

Figure S3 Images showing expression and purification of the binding domains of PAP1 and TT8 induced in *E. coli in vitro*

Figure S4 Promoter sequences of *NtPMT2* and *NtODC* and identification of cis-elements.

Figure S5 Electrophoretic mobility shift assays (EMSAs) showing weak binding of TT8 to G-box-like probe and no binding of PAP1 to MRE-like probe.

Figure S6 Dual-luciferase and Chip-qPCR experiments showing no activation of *NtPMT2* and *NtODC2* promoters by PAP1 and TT8 alone, PAP1 and TT8 together, and PAP1-TT8-WD complex.

Figure S7 Diagrams showing gene stacking design for DNA synthesis, binary construct, and selection of T2 homozygous plants.

Figure S8 Comparison of *AtPAP1* and *AtTT8 trans*gene expressions in roots of wild-type Narrow Leaf Madole (NL) and KY171, P+T transgenic NL and KY171, and vector control transgenic plants.

Figure S9 Phenotypes of wild type, vector-control and P+T-NL and P+T-KY tobacco plants used for analysis of gene expression and metabolites.

Figure S10 Transcriptional upregulation of eight flavonoid pathway genes in red transgenic P+T-NL and P+T-KY tobacco plants.

Figure S11 Anthocyanin levels in roots and leaves of wild type and red P+T transgenic tobacco plants.

Figure S12 Proanthocyanidin formation in roots of red P+T transgenic tobacco plants

Figure S13 Comparison of *NtJAZ7* and *NtJAZ10* gene expression in roots of wild-type Narrow Leaf Madole (NL) and KY171, P+T transgenic NL and KY171, and vector control transgenic plants.

Figure S14 Comparison of *MYC1a*, *MYC1b*, *MYC2a* and *MYC2b* gene expression in roots of wild-type (WT) Narrow Leaf Madole (NL) and KY171, P+T transgenic NL and KY, and vector control transgenic plants.

Figure S15 Schematic diagram showing procedures for field farming practice and leaf harvest from wild type and red transgenic tobacco plants.

Figure S16 Field design for farming practice of red P+T transgenic plants.

Figure S17 Planting, plant growth, and phenotypes of wild type and red P+T transgenic tobacco plants in field from the first day to 75 days after planting.

Figure S18 Harvest of plants from field for air-curing and cleanup of field.
Figure S19 Comparison of leaf color after air-curing and grouping of leaves for sampling.
Figure S20 Reduction of total tobacco alkaloids in each leaf group of red P+T-NL1 and P+T-KY1 plants compared to wild type ones.
Figure S21 Upregulation of *NtAn1a* and *NtAn1b* in roots of red PAP1 tobacco plants.
Figure S22 Field farming practice of PAP1 tobacco plants and leaf harvest
Figure S23 Reduction of nicotine, nornicotine, anabasine, and anatabine in PAP1 tobacco leaves
Figure S24 Reduction of total nicotine, nornicotine, anabasine, and anatabine in all leaves of PAP1 tobacco plant.
Figure S25 Reduction of NNN, NNK, NAB, and NAT in flue-cured leaves of PAP1 tobacco compared to leaves of wild type Xanthi control
Figure S26 Reduction of total TSNAs in flue-cured leaves of PAP1 tobacco.

9 Supporting Table capitals

Table S1 Thermal cycle and sequences of primers designed for PCR, qRT-PCR, Q-PCR, respectively

Table S2 Float trays used for seedlings in the water nursery bed in the greenhouse. The same trays were prepared for PAP1 tobacco plants

Table S3 Instrument Parameters used in Tobacco Specific Nitrosamines analysis

Table S4 HPLC Gradient used in Tobacco Specific Nitrosamines analysis

Table S5 Typical Parameter Table Triple Quadrupole Mass Spectrometer used in Tobacco Specific Nitrosamines analysis

Table S6 Reagents and Standards in for Tobacco Specific Nitrosamines analysis

Table S7 Instrument Parameters used in Tobacco Alkaloids analysis

Table S8 Oven Program used for in Tobacco Alkaloids analysis

Table S9 Reagents and Standards used for Tobacco Alkaloids analysis

References

Andersen, R. A., Fleming, P. D., Burton, H. R., Hamilton-Kemp, T. R., Hildebrand, D. F., Sutton, T. G., 1990 Levels of alkaloids and their derivatives in air- and fire-cured KY 171 dark tobacco during prolonged storage: effects of temperature and moisture. Tobacco Int 192 49-55

Arany, I., Taylor, M., Fulop, T., Dixit, M., 2018. Adverse effects of chronic nicotine exposure on the kidney: Potential human health implications of experimental findings. International Journal of Clinical Pharmacology and Therapeutics 56, 501-506.

Bai, Y. H., Pattanaik, S., Patra, B., Werkman, J. R., Xie, C. H., Yuan, L., 2011. Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active. Planta 234, 363-375.

Baldwin, I. T., Zhang, Z. P., Diab, N., Ohnmeiss, T. E., McCloud, E. S., Lynds, G. Y., Schmelz, E. A., 1997. Quantification, correlations and manipulations of wound-induced changes in jasmonic acid and nicotine in *Nicotiana sylvestris*. Planta 201, 397-404.

Baudry, A., Caboche, M., Lepiniec, L., 2006. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J 46, 768-779.

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., Lepiniec, L., 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 39, 366-380.

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., Lamb, C., 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12, 2383-2394.

Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., Schijlen, E. G. W. M., Hall, R. D., Bovy, A. G., Luo, J., Martin, C., 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotech 26, 1301-1308.

Chakrabarti, M., Bowen, S. W., Coleman, N. P., Meekins, K. M., Dewey, R. E., Siminszky, B., 2008. CYP82E4-mediated nicotine to nornicotine conversion in tobacco is regulated by a senescence-specific signaling pathway. Plant Mol Biol 66, 415-427.

Chini, A., Fonseca, S., Fernandez, G., Adie, B., Chico, J. M., Lorenzo, O., Garcia-Casado, G., Lopez-Vidriero, I., Lozano, F. M., Ponce, M. R., Micol, J. L., Solano, R., 2007. The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448, 666-+.

Chintapakorn, Y., Hamill, J. D., 2003. Antisense-mediated down-regulation of putrescine N-methyltransferase activity in transgenic *Nicotiana tabacum* L.can lead to elevated levels of anatabine at the expense of nicotine. Plant Molecular Biology 53, 87-105.

Dawson, R. F., 1945. An experimental analysis of alkaloid production in *Nicotiana* - The origin of nornicotine. Am J Bot 32, 416-423.

De Boer, K., Tilleman, S., Pauwels, L., Vanden Bossche, R., De Sutter, V., Vanderhaeghen, R., Hilson, P., Hamill, J. D., Goossens, A., 2011. APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J 66, 1053-1065.

1232 DeBoer, K. D., Dalton, H. L., Edward, F. J., Hamill, J. D., 2011. RNAi-mediated down-regulation
1233 of ornithine decarboxylase (ODC) leads to reduced nicotine and increased anatabine levels in
1234 transgenic *Nicotiana tabacum* L. Phytochemistry 72, 344-355.
1235 DeBoer, K. D., Dalton, H. L., Edward, F. J., Ryan, S. M., Hamill, J. D., 2013. RNAi-mediated
1236 down-regulation of ornithine decarboxylase (ODC) impedes wound-stress stimulation of
1237 anabasine synthesis in *Nicotiana glauca*. Phytochemistry 86, 21-28.
1238 DeBoer, K. D., Lye, J. C., Aitken, C. D., Su, A. K. K., Hamill, J. D., 2009. The A622 gene in
1239 *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis.
1240 Plant Mol Biol 69, 299-312.
1241 Dewey, R. E., Xie, J. H., 2013. Molecular genetics of alkaloid biosynthesis in Nicotiana
1242 tabacum. Phytochemistry 94, 10-27.
1243 FDA, 2017. Tobacco Product Standard for N-Nitrosonornicotine Level in Finished Smokeless
1244 Tobacco Products. In: SERVICES, H. A. H. (Ed.). Food and Drug Administration.
1245 Feng, K., Hou, X. L., Xing, G. M., Liu, J. X., Duan, A. Q., Xu, Z. S., Li, M. Y., Zhuang, J., Xiong, A.
1246 S., 2020. Advances in AP2/ERF super-family transcription factors in plant. Crit. Rev.
1247 Biotechnol. 40, 750-776.
1248 Gatica-Arias, A., Farag, M. A., Stanke, M., Matousek, J., Wessjohann, L., Weber, G., 2012.
1249 Flavonoid production in transgenic hop (*Humulus lupulus* L.) altered by PAP1/MYB75 from
1250 *Arabidopsis thaliana* L. Plant Cell Rep 31, 111-119.
1251 Gavilano, L. B., Siminszky, B., 2007. Isolation and characterization of the cytochrome P450
1252 gene CYP82E5v2 that mediates nicotine to nornicotine conversion in the green leaves of
1253 tobacco. Plant Cell Physiol 48, 1567-1574.
1254 Gonzalez, A., Zhao, M., Leavitt, J. M., Lloyd, A. M., 2008. Regulation of the anthocyanin
1255 biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis*
1256 seedlings. Plant J 53, 814-827.
1257 Goossens, J., Mertens, J., Goossens, A., 2017. Role and functioning of bHLH transcription
1258 factors in jasmonate signalling. J Exp Bot 68, 1333-1347.
1259 Grando, S. A., 2014. Connections of nicotine to cancer. Nature Reviews Cancer 14, 419-429.
1260 Greillier, L., Cortot, A. B., Viguier, J., Brignoli-Guibaudet, L., Lhomel, C., Eisinger, F., Morere, J.
1261 F., Couraud, S., 2018. Perception of Lung Cancer Risk: Impact of Smoking Status and Nicotine
1262 Dependence. Current Oncology Reports 20, 7.
1263 He, X., Li, Y., Lawson, D., Xie, D.-Y., 2017. Metabolic engineering of anthocyanins in dark
1264 tobacco varieties Physiol Plant 159, 2-12.
1265 Hecht, S. S., 1998. Biochemistry, biology, and carcinogenicity of tobacco-specific N-
1266 nitrosamines. Chemical Research in Toxicology 11, 559-603.
1267 Hecht, S. S., Stepanov, I., Carmella, S. G., 2016. Exposure and Metabolic Activation
1268 Biomarkers of Carcinogenic Tobacco-Specific Nitrosamines. Accounts of Chemical Research
1269 49, 106-114.
1270 Hibi, N., Higashiguchi, S., Hashimoto, T., Yamada, Y., 1994. Gene-expression in tobacco low-
1271 nicotine mutants. Plant Cell 6, 723-735.
1272 Howe, G. A., Major, I. T., Koo, A. J., 2018. Modularity in Jasmonate Signaling for Multistress
1273 Resilience. In: Merchant, S. S. (Ed.), Annual Review of Plant Biology, Vol 69, vol. 69. Annual
1274 Reviews, Palo Alto, pp. 387-415.

1275 Imanishi, S., Hashizume, K., Nakakita, M., Kojima, H., Matsubayashi, Y., Hashimoto, T.,
1276 Sakagami, Y., Yamada, Y., Nakamura, K., 1998. Differential induction by methyl jasmonate of
1277 genes encoding ornithine decarboxylase and other enzymes involved in nicotine biosynthesis
1278 in tobacco cell cultures. Plant Mol Biol 38, 1101-1111.
1279 Kajikawa, M., Hirai, N., Hashimoto, T., 2009. A PIP-family protein is required for biosynthesis
1280 of tobacco alkaloids. Plant Mol Biol 69, 287-298.
1281 Kajikawa, M., Shoji, T., Kato, A., Hashimoto, T., 2011. Vacuole-localized berberine bridge
1282 enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant
1283 Physiol 155, 2010-2022.
1284 Konstantinou, E., Fotopoulou, F., Drosos, A., Dimakopoulou, N., Zagoriti, Z., Niarchos, A.,
1285 Makrynioti, D., Kouretas, D., Farsalinos, K., Lagoumintzis, G., Poulas, K., 2018. Tobacco-
1286 specific nitrosamines: A literature review. Food Chem. Toxicol. 118, 198-203.
1287 Ladesic, B., Tso, T. C., 1964. Biochemical studies on tobacco alkaloids. 6. Biosynthesis of
1288 nicotine through nornicotine. Phytochemistry 3, 541-545.
1289 Lee, P. N., Forey, B. A., Hamling, J. S., Thornton, A. J., 2017. Environmental tobacco smoke
1290 exposure and heart disease: A systematic review. World Journal of Meta-Analysis 5, 14-40.
1291 Lewis, R. S., Bowen, S. W., Keogh, M. R., Dewey, R. E., 2010. Three nicotine demethylase
1292 genes mediate nornicotine biosynthesis in *Nicotiana tabacum* L.: Functional characterization
1293 of the CYP82E10 gene. Phytochemistry 71, 1988-1998.
1294 Lewis, R. S., Jack, A. M., Morris, J. W., Robert, V. J. M., Gavilano, L. B., Siminszky, B., Bush, L.
1295 P., Hayes, A. J., Dewey, R. E., 2008. RNA interference (RNAi)-induced suppression of nicotine
1296 demethylase activity reduces levels of a key carcinogen in cured tobacco leaves. Plant
1297 Biotechnol J 6, 346-354.
1298 Lewis, R. S., Lopez, H. O., Bowen, S. W., Andres, K. R., Steede, W. T., Dewey, R. E., 2015.
1299 Transgenic and mutation-based suppression of a berberine bridge enzyme-like (BBL) gene
1300 family reduces alkaloid content in field-grown tobacco. PLoS One 10.
1301 Li, M. Z., Li, Y. Z., Guo, L. L., Gong, N. D., Pang, Y. Z., Jiang, W. B., Liu, Y. J., Jiang, X. L., Zhao, L.,
1302 Wang, Y. S., Xie, D. Y., Gao, L. P., Xia, T., 2017. Functional characterization of tea (*Camellia*
1303 *sinensis*) MYB4a transcription factor using an integrative approach. Front. Plant Sci. 8, 17.
1304 Li, X., Gao, M. J., Pan, H. Y., Cui, D. J., Gruber, M. Y., 2010. Purple canola: Arabidopsis PAP1
1305 increases antioxidants and phenolics in *Brassica napus* leaves. J Agric Food Chem 58, 1639-
1306 1645.
1307 Millard, P. S., Kragelund, B. B., Burow, M., 2019. R2R3 MYB Transcription Factors - Functions
1308 outside the DNA-Binding Domain. Trends Plant Sci. 24, 934-946.
1309 Mizusaki, S., Tanabe, Y., Noguchi, M., Tamaki, E., 1973. Phytochemical studies on tobacco
1310 alkaloids. 16. Changes in activities of ornitheine decarboxylase, putrescine N-
1311 methyltransferase and N-methyl-putrescine oxidase in tobacco roots in relation to nicotine
1312 biosynthesis. Plant Cell Physiol 14, 103-110.
1313 Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., Lepiniec, L., 2000. The *TT8* gene
1314 encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN*
1315 genes in *Arabidopsis* siliques. The Plant Cell 12, 1863-1878.

1316 Pearce, B., Miller, B., Walker, E., Vann, M., Whitley, S., 2015. Selecting Burley Tobacco
1317 Varieties. In: Pearce, B., Bailey, A., Walker, E. (Eds.), 2015-2016 Burley and Dark Tobacco
1318 Production Guide, pp3-7. University of Kentucky, Lexington.
1319 Pireyre, M., Burow, M., 2015. Regulation of MYB and bHLH Transcription Factors: A Glance
1320 at the Protein Level. Mol. Plant. 8, 378-388.
1321 Qi, T., Song, S., Ren, Q., Wu, D., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C., Xie, D., 2011.
1322 The jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to
1323 regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in
1324 *Arabidopsis thaliana*. Plant Cell 23, 1795-1814.
1325 Qi, T. C., Huang, H., Wu, D. W., Yan, J. B., Qi, Y. J., Song, S. S., Xie, D. X., 2014. Arabidopsis
1326 DELLA and JAZ proteins bind the WD-Repeat/bHLH/MYB complex to modulate gibberellin
1327 and jasmonate signaling synergy. Plant Cell 26, 1118-1133.
1328 Ramsay, N. A., Glover, B. J., 2005. MYB-bHLH-WD40 protein complex and the evolution of
1329 cellular diversity. Trends Plant Sci 10, 63-70.
1330 Saigo, T., Wang, T., Watanabe, M., Tohge, T., 2020. Diversity of anthocyanin and
1331 proanthocyanin biosynthesis in land plants. Curr. Opin. Plant Biol. 55, 93-99.
1332 Sanner, T., Grimsrud, T. K., 2015. Nicotine: carcinogenicity and effects on response to cancer
1333 treatment - a review. Frontiers in Oncology 5, 10.
1334 Santoro, A., Tomino, C., Prinzi, G., Lamonaca, P., Cardaci, V., Fini, M., Russo, P., 2019.
1335 Tobacco Smoking: Risk to Develop Addiction, Chronic Obstructive Pulmonary Disease, and
1336 Lung Cancer. Recent Patents on Anti-Cancer Drug Discovery 14, 39-52.
1337 Shi, M.-Z., Xie, D.-Y., 2014. Biosynthesis and metabolic engineering of anthocyanins in
1338 *Arabidopsis thaliana*. Recent Pat Biotechnol 8, 47-60.
1339 Shi, M. Z., Xie, D. Y., 2011. Engineering of red cells of *Arabidopsis thaliana* and comparative
1340 genome-wide gene expression analysis of red cells versus wild-type cells. Planta 233, 787-
1341 805.
1342 Shoji, T., Hashimoto, T., 2011a. Recruitment of a duplicated primary metabolism gene into
1343 the nicotine biosynthesis regulon in tobacco. Plant J 67, 949-959.
1344 Shoji, T., Hashimoto, T., 2011b. Tobacco MYC2 regulates jasmonate-inducible nicotine
1345 biosynthesis genes directly and by way of the NIC2-locus ERF genes. Plant Cell Physiol 52,
1346 1117-1130.
1347 Shoji, T., Hashimoto, T., 2013. Smoking out the masters: transcriptional regulators for
1348 nicotine biosynthesis in tobacco. Plant Biotechnol 30, 217-224.
1349 Shoji, T., Ogawa, T., Hashimoto, T., 2008. Jasmonate-induced nicotine formation in tobacco
1350 is mediated by tobacco COI1 and JAZ genes. Plant Cell Physiol 49, 1003-1012.
1351 Shoji, T., Yamada, Y., Hashimoto, T., 2000. Jasmonate induction of putrescine N-
1352 methyltransferase genes in the root of *Nicotiana sylvestris*. Plant Cell Physiol 41, 831-839.
1353 Skinner, M. K., Rawls, A., Wilson-Rawls, J., Roalson, E. H., 2010. Basic helix-loop-helix
1354 transcription factor gene family phylogenetics and nomenclature. Differentiation 80, 1-8.
1355 Springob, K., Nakajima, H., Yamazaki, M., Saito, K., 2003. Recent advances in the biosynthesis
1356 and accumulation of anthocyanins. Nat Prod Rep 20, 288-303.
1357 Stepka, W., Dewey, L. J., 1961. Conversion of nicotine to nornicotine in harvested tobacco -
1358 fate of methyl group. Plant Physiol 36, 592-&.

1359 Thevenin, J., Dubos, C., Xu, W., Le Gourrierec, J., Kelemen, Z., Charlot, F., Nogue, F., Lepiniec,
1360 L., Dubreucq, B., 2012. A new system for fast and quantitative analysis of heterologous gene
1361 expression in plants. New Phytol. 193, 504-512.
1362 Thorndike, A. N., Rigotti, N. A., 2009. A tragic triad: coronary artery disease, nicotine
1363 addiction, and depression. Current Opinion in Cardiology 24, 447-453.
1364 Tidey, J. W., Davis, D. R., Miller, M. E., Pericot-Valverde, I., Denlinger-Apte, R. L., Gaalema, D.
1365 E., 2018. Modeling nicotine regulation: A review of studies in smokers with mental health
1366 conditions. Prev. Med. 117, 30-37.
1367 Tso, T. C., Jeffrey, R. N., 1956. Studies on tobacco alkaloids. 1. Changes in nicotine and
1368 nornicotine content in *Nicotiana*. Plant Physiol 31, 433-439.
1369 Tso, T. C., Jeffrey, R. N., 1957. Studies on tobacco alkaloids. 2. The formation of nicotine and
1370 nornicotine in tobacco supplied with N15. Plant Physiol 32, 86-92.
1371 Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasdan, N.,
1372 Blundell, T. L., Esch, J. J., Marks, M. D., Gray, J. C., 1999. The *TRANSPARENT TESTA GLABRA 1*
1373 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis,
1374 encodes a WD40 repeat protein. Plant Cell 11, 1377-1349.
1375 Wang, B. W., Lewis, R. S., Shi, J. L., Song, Z. B., Gao, Y. L., Li, W. Z., Chen, H. X., Qu, R. D.,
1376 2015. Genetic factors for enhancement of nicotine levels in cultivated tobacco. Sci Rep 5.
1377 Wang, P., Liu, Y., Zhang, L., Wang, W., Hou, H., Zhao, Y., Jiang, X., Yu, J., Tan, H., Wang, Y.,
1378 Xie, D.-Y., Gao, L., Xia, T., 2020. Functional demonstration of plant flavonoid carbocations
1379 proposed to be involved in the biosynthesis of proanthocyanidins. The Plant Journal 101, 18-
1380 36.
1381 Xie, D.-Y., Dixon, R. A., 2005. Proanthocyanidin biosynthesis - still more questions than
1382 answers? Phytochemistry 66, 2127-2144.
1383 Xie, D.-Y., Sharma, S. B., Paiva, N. L., Ferreira, D., Dixon, R. A., 2003. Role of anthocyanidin
1384 reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science 299, 396-399.
1385 Xie, D.-Y., Sharma, S. B., Wright, E., Wang, Z.-Y., Dixon, R. A., 2006. Metabolic engineering of
1386 proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB
1387 transcription factor. Plant J 45, 895-907.
1388 Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M., Turner, J. G., 1998. COI1: An Arabidopsis
1389 gene required for jasmonate-regulated defense and fertility. Science 280, 1091-1094.
1390 Xie, J., Song, W., Maksymowicz, W., Jin, W., Cheah, K., Chen, W., Carnes, C., Ke, J., Conkling,
1391 M. A., 2004a. Biotechnology: A tool for reduced-risk tobacco products - The nicotine
1392 experience from test tube to cigarette pack. RECENT ADVANCES IN TOBACCO SCIENCE, vol.
1393 30, pp. 17-38.
1394 Xie, J., W, S., Maksymowicz, W., W, J., K, C., Wx, C., C, C., J, K., Ma, C., 2004b. Biotechnology:
1395 A tool for reduced risk tobacco products - The nicotine experience from test tube to
1396 cigarette pack. Recent Advances in Tobacco Science, vol. 30, pp. 17-37.
1397 Xu, B. F., Timko, M. P., 2004. Methyl jasmonate induced expression of the tobacco
1398 putrescine N-methyltransferase genes requires both G-box and GCC-motif elements. Plant
1399 Mol Biol 55, 743-761.
1400 Xu, W. J., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., Lepiniec, L., Dubos,
1401 C., 2014. Complexity and robustness of the flavonoid transcriptional regulatory network

revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. New Phytol. 202, 132-144.
Xu, W. J., Grain, D., Le Gourrierec, J., Harscoet, E., Berger, A., Jauvion, V., Scagnelli, A., Berger, N., Bidzinski, P., Kelemen, Z., Salsac, F., Baudry, A., Routaboul, J. M., Lepiniec, L., Dubos, C., 2013. Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression, in Arabidopsis. New Phytol. 198, 59-70.
Zhang, H. B., Bokowiec, M. T., Rushton, P. J., Han, S. C., Timko, M. P., 2012. Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. Mol Plant 5, 73-84.
Zhang, H. Y., Li, W. J., Niu, D. X., Wang, Z. J., Yan, X. X., Yang, X. L., Yang, Y. F., Cui, H., 2019. Tobacco transcription repressors NtJAZ: Potential involvement in abiotic stress response and glandular trichome induction. Plant Physiol. Biochem. 141, 388-397.
Zhao, D., Qin, L. J., Zhao, D. G., 2016. RNA interference of the nicotine demethylase gene CYP82E4v1 reduces nornicotine content and enhances *Myzus persicae* resistance in *Nicotiana tabacum* L. Plant Physiol Biochem 107, 214-221.
Zhu, Z. X., Wang, H. L., Wang, Y. T., Guan, S., Wang, F., Tang, J. Y., Zhang, R. J., Xie, L. L., Lu, Y. Q., 2015. Characterization of the cis elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. J. Exp. Bot. 66, 3775-3789.
Zuluaga, D. L., Gonzali, S., Loreti, E., Pucciariello, C., Degl'Innocenti, E. D., Guidi, L., Alpi, A., Perata, P., 2008. Arabidopsis thaliana MYB75/PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. Funct Plant Biol 35, 606-618.
Zvi, M. M. B., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S., Ovadis, M., Vainstein, A., 2012. PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. New Phytol., 10.1111/j.1469-8137.2012.04161.x.

Figure Legends

Figure 1 Biosynthetic pathways of nicotine, other tobacco alkaloids, and anthocyanin and their regulations. A, the biosynthetic pathway of nicotine, nornicotine, and tobacco specific nitrosamines (TSNAs) and the regulation mechanism. Inserted two photos show wild type Xanthi plants grown in the field and root color. MYC2 and EAR are two positive transcription factors that activate the expression of most pathway genes. In the absence of isoleucine-jasmonate (Ile-JA), NtJAZ 1 and NtJAZ3 repressors bind to MYC2 to inhibit the activation of the pathway genes (red lines). In the presence of bioactive Ile-JA (green arrows), SCF^{COI1} perceives this plant hormone to form a new JA-SCF^{COI1} complex to bind to NtJAZs and lead NtJAZs to proteasomal ubiquitination. The degradation of NtJAZs releases

MYC2 to activate pathway gene expression toward the biosynthesis of nicotine. Gene abbreviations are *AO*: aspartate oxidase, *QS*: quinolinate synthase, *QPT2*: quinolinate phosphoribosyltransferase 2, *ADC:* arginine decarboxylase, *AIH*: agmatine iminohydrolase, *ODC*: ornithine decarboxylase, *MPT1*: N-methylputricine transferase 1, *MPO1*: methylputricine oxidase 1, *A622*: an isoflavone reductase like enzyme, *BBL*: berberine bridge enzyme-like protein, and *NND*: nicotine N-demethylase. **B,** structures of nicotine, nornicotine and four TSNAs. **C**, the biosynthetic pathway of Arabidopsis anthocyanin from the step catalyzed by CHS. *PAP1*, *TT8*, and *TTG1* encode MYB75, bHLH, and WD40 to form a MYB-bHLH-WD40 complex that activates late pathway genes. Gene abbreviations are *CHS* and *CHI*: chalcone synthase and isomerase, *F3H*: flavonone-3-hydroxylase, *F3'H* and *F3'5'H*: flavonoid-3'-hydroxylase and 3'5'-hydroxylase, *DFR*: dihydroflavonol reductase, *ANS*: anthocyanidin synthase, *3-GT*: glycosyltransferase. *ANR* encodes anthocyanidin reductase, a key enzyme toward proanthocyanidin biosynthesis. Inserted two photos show a novel red genotypic PAP1 tobacco (programmed by Arabidopsis PAP1) grown in the field and dark root color.

Figure 2 Binding and activation of *NtJAZ1* and *NtJAZ3* promoters by PAP1 and TT8 alone, PAP1 and TT8 together, and PAP1-TT8-WD40 complex. A, three MYB response element (MRE) types and G-box were identified in the *NtJAZ1* and *NtJAZ3* promoters; **B**, electrophoretic mobility shift assay (EMSA) showed that PAP1 and TT8 bound to three MRE types and G-box of the *NtJAZ1* and *NtJAZ3* promoters, respectively. Competitive and non-competitive probes were used for binding of both TFs. 20× and 10×: concentrations of tested competitive and non-competitive probes 20 and 10 times as those MRE and G-box probes. **C**, fold change values from Chip-qPCR showed that both PAP1 and TT8 bound to the *NtJAZ1* and *NtJAZ3* promoters *in vivo*. Approximately 150 bp *NtJAZ1* and *NtJAZ3* promoter fragments containing both MRE and G-box were enriched with anti-HA antibodies in qRT-PCR analysis. The regions of *NtJAZ1* and *NtJAZ3* promoters that do not contain a MRE and G-box were used as negative controls. Green and red bars represent wild type and red tobacco plants, respectively. **D**, schematic diagrams shows four effector constructs (PK2GW7-GFP, PAP1, TT8 and TTG1) and two reporter constructs (pGreenII-0800-Jaz1pro and Jaz3 pro) used for dual-luciferase assay. **E**, luciferase (LUC)/renilla (REN) luminescence ratios from dual luciferase assays showed that PAP1 and TT8 alone bound to and activated the promoters of both *NtJAZ1* and *NtJAZ3*, two TF together increased promoter activity, and two TFs and WD40 together increased the highest activity. The promoters were fused to a firefly luciferase (reporter). The promoters of *NtJAZ1* and *NtJAZ3* genes were used in dual luciferase assay. Values represent the mean ± S.D. (n=5). Asterisks on top of bars indicate that the values are

significantly higher than those of controls (*P<0.05, **P<0.01, ***P<0.001). Standard bars mean standard errors.

Figure 3 Upregulation of two *NtJAZs* and downregulation of six or seven nicotine pathway genes and *ERF189* in roots of P+T transgenic seedlings shown by qRT-PCR

A-J, in roots of P+T-NL transgenic lines, the expression levels of *NtJAZ1* (**A**) and *NtJAZ3* (**B**) were significantly upregulated; the expression level of *ODC2* (**C**) was not changed; and the expression levels of *PMT1* (**D**), *PMT2* (**E**) *MPO* (**F**), *QPT* (**G**), *A622* (**H**), *BBLs* (**I**) **(**primer pairs designed for all three BBLs), and *ERF189* (**J**) were significantly downregulated. **K-T**, in roots of P+T-KY transgenic lines, the expression levels of *NtJAZ1* (**K**) and *NtJAZ3* (**L**) were significantly upregulated, and the expression levels *ODC2* (**M**), *PMT1* (**N**), *PMT2* (**O**) *MPO* (**P**), *QPT* (**Q**), *A622* (**R**), *BBLs* (**S**), and *ERF189* (**T**) were significantly downregulated. Data of three transgenic lines are shown for both P+T-NL and P+T-KY genotypes. Wild type samples were pooled from five plants. Data of three lines are shown for vector control transgenic NL and KY171 plants. WT: wild type, P+T NL1, P+T NL2, P+T NL3: three lines of stacked *PAP1* and *TT8* transgenic NL plants, VC-1, 2, and 3-NL: three vector control transgenic NL lines, P+T KY1, P+T KY2, P+T KY3: three lines of stacked *PAP1* and *TT8* transgenic KY171 plants. VC-1, 2, and 3-KY: three vector control transgenic KY171 lines.

Green and red bars represent wild type and vector control transgenic, and red P+T-NL and P+T-KY plants, respectively. Values represent the mean ± S.D. (n=3). Asterisks on top of the bars indicate that values are significantly lower or higher in red transgenic lines than in wild type plants (*P<0.05, **P<0.01, ***P<0.001). Standard bars mean standard errors.

Figure 4 Significant reduction of nicotine and nornicotine contents in leaves and roots of P+T-NL and P+T-KY seedlings. The contents of nicotine and nornicotine were significantly reduced in roots and leaves of both transgenic P+T-NL and P+T-KY lines. **A-C:** reduction of nicotine (**A**), nornicotine (**B**), and total nicotine and nornicotine (**C**) in leaves of three P+T-NL lines (1, 2, and 3); **D-F:** reduction of nicotine (**D**), nornicotine (**E**), and total nicotine and nornicotine (**F**) in roots of three P+T-NL lines (1, 2, and 3); **G-I:** reduction of nicotine (**G**), nornicotine (**H**), and total nicotine and nornicotine (**I**) in leaves of three P+T-KY lines (1, 2 and 3); **J-L:** reduction of nicotine (**J**), nornicotine (**K**), and total nicotine and nornicotine (**L**) in roots of three P+T-NL lines (1, 2, and 3). Bars labeled with “a” and “b” means p-

value less than 0.05 and bars labelled with the same low case “a” or “b” means no significant difference.

Figure 5 Phenotypes of P+T-NL1 and P+T-KY1 plants versus their corresponding wild-type tobacco plants in the field and of air-cured leaves. Field farming practice of four genotypes was performed in Oxford, North Carolina. **A**, phenotypes of 30-day and 60-day old WT-NL vs. red P+T-NL1 plants. **B**, phenotypes of 30-day and 60-day old WT-KY171 vs. red P+T-KY1 plants. **C-D**, phenotypes of topped plants (**C**) and leaves in air curing (**D**). Plant name abbreviations are WT-NL: wild-type Narrow Leaf Madole (NL) variety, P+T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WT-KY171: wild-type KY171 variety, and P+T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.

Figure 6 Reduction of nicotine and four other alkaloids in cured leaves of red P+T-NL1 and P+T-KY1 tobacco plants. A-J, reduction of nicotine (**A-B**), nornicotine (**C-D**), anabasine (**E-F**), anatabine (**G-H**), and myosmine (**I-J**) in four leaf groups of P+T-NL (**A, C, E, G, and I**) and P+T-KY plants (**B, D, F, H, and J**). **K-L**, cartoons showing leaf groups harvested from both wild type (**K**) and transgenic red plants (**L**).

Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of tobacco plant, respectively. Each metabolite in each leaf group was quantified from leaves harvested from 120 tobacco plants in three plots. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Bars labelled with low case “a” and “b” mean significant difference between paired green and red bars (n=120, P-value less than 0.05). Percentages labeled in the top of red bars mean reduction compared to green bars. The standard bars mean standard deviation. Abbreviations, WT-NL: wild-type Narrow Leaf Madole (NL) variety, P+T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WT-KY: wild-type KY171 variety, and P+T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.

Figure 7 Reduction of four individual and total TSNAs in leaves of red P+T-NL1 and P+T-KY1 tobacco plants. A-H, reduction of NNN (**A-B**), NNK (**C-D**), NAT (**E-F**), and NAB (**G-H**) in each leaf group of P+T-NL1 (**A, C, E,** and **G**) and P+T-KY1 (**B, D, F,** and **H**) plants; **I-J,** reduction of total TSNA contents summed from

four individual TSNAs of each leaf group of P+T-NL1 (**I**) and P+T-KY1 (**J**) plants; **K-L,** reduction of total TSNA contents summed from four individual TSNAs of all leaves of P+T-NL1 (**K**) and P+T-KY1 (**L**) plants.

Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of plants, respectively. Each metabolite in each group of leaves was quantified from 120 tobacco plants. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Paired green and red bars labelled with low case “a” and “b” mean significant difference (n=120, P-value less than 0.05). Percentages labeled on the top of red bars indicate reduction levels compared to green bars. Abbreviation, WT-NL: Narrow Leaf Madole, WT-KY171: wild type KY171, P+T-NL: stacked PAP1 and TT8 transgenic NL line 1, and P+T-KY1: stacked PAP1 and TT8 transgenic KY171 line 1.

Figures

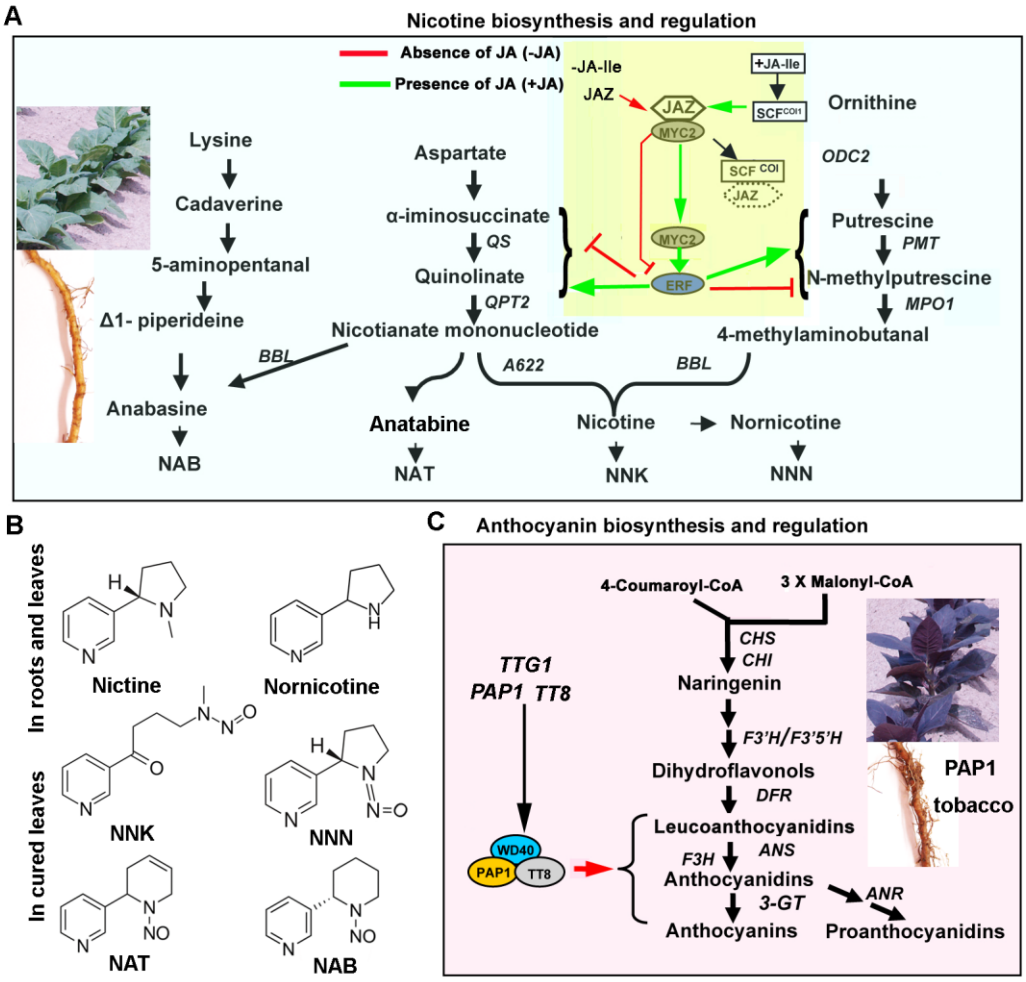
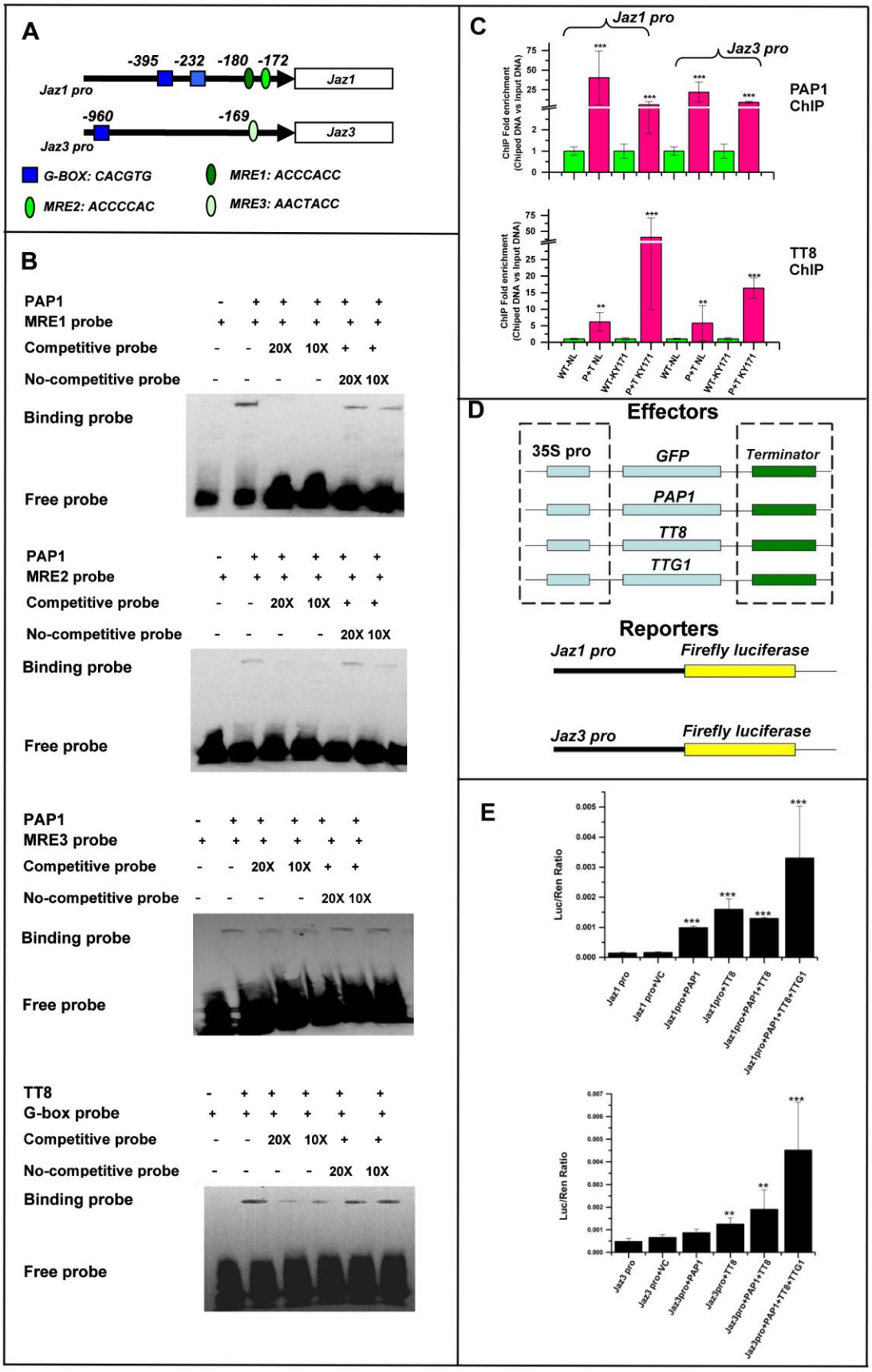


Figure 1 Biosynthetic pathways of nicotine, other tobacco alkaloids, and anthocyanin and their regulations. A, the biosynthetic pathway of nicotine, nornicotine, and tobacco specific nitrosamines (TSNAs) and the regulation mechanism. Inserted two photos show wild type Xanthi plants grown in the field and root color. MYC2 and EAR are two positive transcription factors that activate the expression of most pathway genes. In the absence of isoleucine-jasmonate (Ile-JA), NtJAZ 1 and NtJAZ3 repressors bind to MYC2 to inhibit the activation of the pathway genes (red lines). In the presence of bioactive Ile-JA (green arrows), SCF^{COI1} perceives this plant hormone to form a new JA-SCF^{COI1} complex to bind to NtJAZs and lead NtJAZs to proteasomal ubiquitination. The degradation of NtJAZs releases MYC2 to activate pathway gene expression toward the biosynthesis of nicotine. Gene abbreviations are *AO*: aspartate oxidase, *QS*: quinolinate synthase, *QPT2*: quinolinate phosphoribosyltransferase 2, *ADC:* arginine decarboxylase, *AIH*: agmatine iminohydrolase, *ODC*: ornithine decarboxylase, *MPT1*: N-methylputricine transferase 1, *MPO1*: methylputricine oxidase 1, *A622*: an isoflavone reductase like enzyme, *BBL*: berberine bridge enzyme-like protein, and *NND*: nicotine N-demethylase. **B,**

structures of nicotine, nornicotine and four TSNAs. **C,** the biosynthetic pathway of Arabidopsis anthocyanin from the step catalyzed by CHS. *PAP1*, *TT8*, and *TTG1* encode MYB75, bHLH, and WD40 to form a MYB-bHLH-WD40 complex that activates late pathway genes. Gene abbreviations are *CHS* and *CHI*: chalcone synthase and isomerase, *F3H*: flavonone-3-hydroxylase, *F3'H* and *F3'5'H*: flavonoid-3'-hydroxylase and 3'5'-hydroxylase, *DFR*: dihydroflavonol reductase, *ANS*: anthocyanidin synthase, *3-GT*: glycosyltransferase. *ANR* encodes anthocyanidin reductase, a key enzyme toward proanthocyanidin biosynthesis. Inserted two photos show a novel red genotypic PAP1 tobacco (programmed by Arabidopsis PAP1) grown in the field and dark root color.


A
-395 -232 -180 -172
Jaz1 pro
Jaz1
-960 -169
Jaz3 pro
Jaz3
G-BOX: CACGTG
MRE1: ACCCACC
MRE2: ACCCCAC
MRE3: AACTACC
B
PAP1
MRE1 probe
Competitive probe
No-competitive probe
Binding probe
Free probe
PAP1
MRE2 probe
Competitive probe
No-competitive probe
Binding probe
Free probe
PAP1
MRE3 probe
Competitive probe
No-competitive probe
Binding probe
Free probe
TT8
G-box probe
Competitive probe
No-competitive probe
Binding probe
Free probe
C
Jaz1 pro
Jaz3 pro
ChIP Fold enrichment (Chiped DNA vs Input DNA)
PAP1 ChIP
TT8 ChIP
WT-NL
P+T NL
WT-KY171
P+T KY171
D
Effectors
35S pro
GFP
Terminator
PAP1
TT8
TTG1
Reporters
Jaz1 pro
Firefly luciferase
Jaz3 pro
Firefly luciferase
E
Luc/Ren Ratio
Jaz1 pro
Jaz1 pro+VC
Jaz1pro+PAP1
Jaz1pro+TT8
Jaz1pro+PAP1+TT8
Jaz1pro+PAP1+TT8+TTG1
Jaz3 pro
Jaz3 pro+VC
Jaz3pro+PAP1
Jaz3pro+TT8
Jaz3pro+PAP1+TT8
Jaz3pro+PAP1+TT8+TTG1

1599

Figure 2 Binding and activation of *NtJAZ1* and *NtJAZ3* promoters by PAP1 and TT8 alone, PAP1 and TT8 together, and PAP1-TT8-WD40 complex. A, three MYB response element (MRE) types and G-box were identified in the *NtJAZ1* and *NtJAZ3* promoters; **B**, electrophoretic mobility shift assay (EMSA) showed that PAP1 and TT8 bound to three MRE types and G-box of the *NtJAZ1* and *NtJAZ3* promoters, respectively. Competitive and non-competitive probes were used for binding of both TFs. 20× and 10×: concentrations of tested competitive and non-competitive probes 20 and 10 times as those MRE and G-box probes. **C**, fold change values from Chip-qPCR showed that both PAP1 and TT8 bound to the *NtJAZ1* and *NtJAZ3* promoters *in vivo*. Approximately 150 bp *NtJAZ1* and *NtJAZ3* promoter fragments containing both MRE and G-box were enriched with anti-HA antibodies in qRT-PCR analysis. The regions of *NtJAZ1* and *NtJAZ3* promoters that do not contain a MRE and G-box were used as negative controls. Green and red bars represent wild type and red tobacco plants, respectively. **D**, schematic diagrams shows four effector constructs (PK2GW7-GFP, PAP1, TT8 and TTG1) and two reporter constructs (pGreenII-0800-Jaz1pro and Jaz3 pro) used for dual-luciferase assay. **E**, luciferase (LUC)/renilla (REN) luminescence ratios from dual luciferase assays showed that PAP1 and TT8 alone bound to and activated the promoters of both *NtJAZ1* and *NtJAZ3*, two TF together increased promoter activity, and two TFs and WD40 together increased the highest activity. The promoters were fused to a firefly luciferase (reporter). The promoters of *NtJAZ1* and *NtJAZ3* genes were used in dual luciferase assay. Values represent the mean ± S.D. (n=5). Asterisks on top of bars indicate that the values are significantly higher than those of controls (*P<0.05, **P<0.01, ***P<0.001). Standard bars mean standard errors.

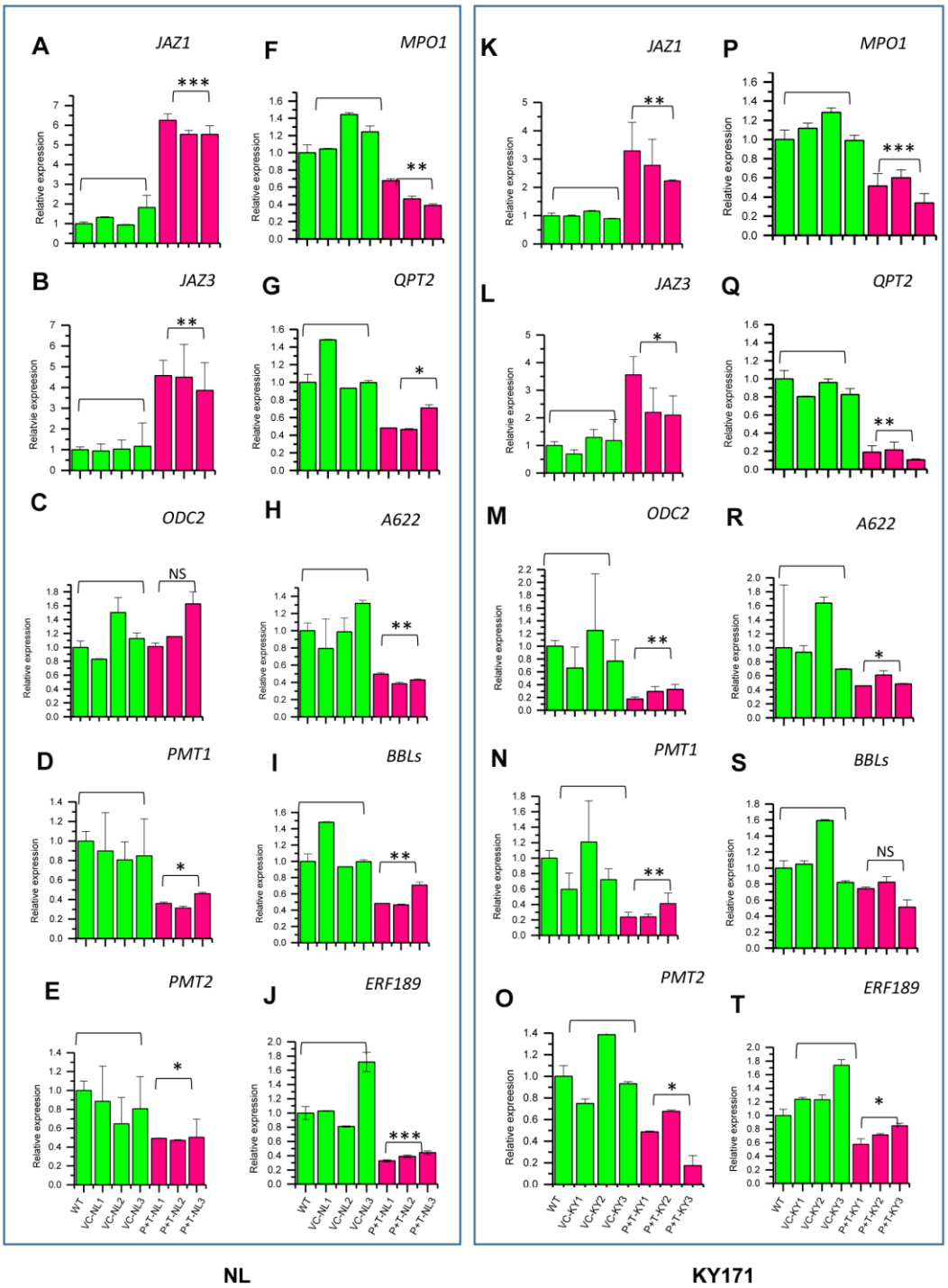


Figure 3 Upregulation of two *NtJAZs* and downregulation of six or seven nicotine pathway genes and *ERF189* in roots of P+T transgenic seedlings shown by qRT-PCR

A-J, in roots of P+T-NL transgenic lines, the expression levels of *NtJAZ1* (**A**) and *NtJAZ3* (**B**) were significantly upregulated; the expression level of *ODC2* (**C**) was not changed; and the expression levels of *PMT1* (**D**), *PMT2* (**E**) *MPO* (**F**), *QPT* (**G**), *A622*

(**H**), *BBLs* (**I**) **(**primer pairs designed for all three BBLs), and *ERF189* (**J**) were significantly downregulated. **K-T**, in roots of P+T-KY transgenic lines, the expression levels of *NtJAZ1* (**K**) and *NtJAZ3* (l) were significantly upregulated, and the expression levels *ODC2* (**M**), *PMT1* (**N**), *PMT2* (**O**) *MPO* (**P**), *QPT* (**Q**), *A622* (**R**), *BBLs* (**S**), and *ERF189* (**T**) were significantly downregulated. Data of three transgenic lines are shown for both P+T-NL and P+T-KY genotypes. Wild type samples were pooled from five plants. Data of three lines are shown for vector control transgenic NL and KY171 plants. WT: wild type, P+T NL1, P+T NL2, P+T NL3: three lines of stacked *PAP1* and *TT8* transgenic NL plants, VC-1, 2, and 3-NL: three vector control transgenic NL lines, P+T KY1, P+T KY2, P+T KY3: three lines of stacked *PAP1* and *TT8* transgenic KY171 plants. VC-1, 2, and 3-KY: three vector control transgenic KY171 lines.

Green and red bars represent wild type and vector control transgenic, and red P+T-NL and P+T-KY plants, respectively. Values represent the mean ± S.D. (n=3). Asterisks on top of the bars indicate that values are significantly lower or higher in red transgenic lines than in wild type plants (*P<0.05, **P<0.01, ***P<0.001). Standard bars mean standard errors.

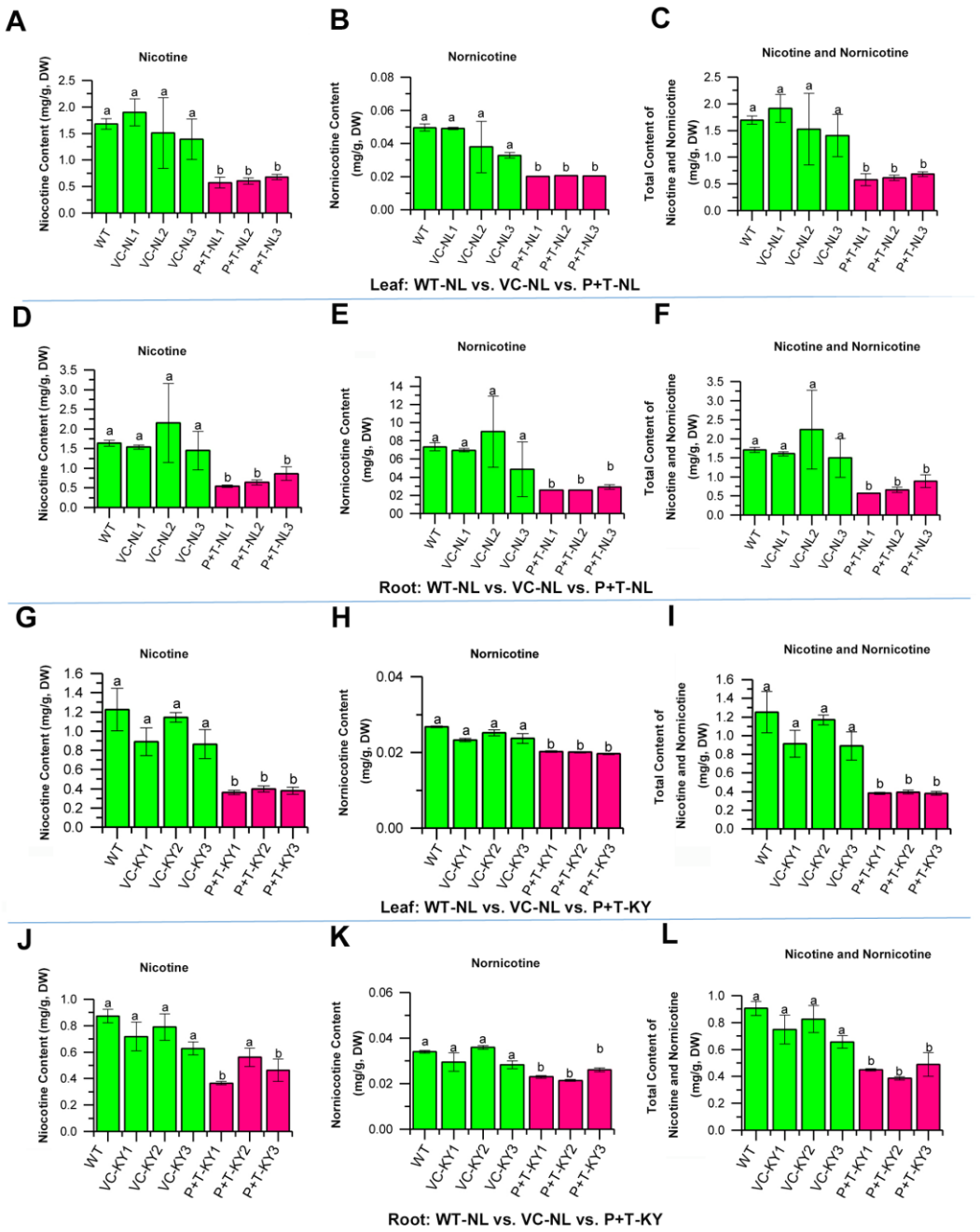


Figure 4 Significant reduction of nicotine and nornicotine contents in leaves and roots of P+T-NL and P+T-KY seedlings. The contents of nicotine and nornicotine were significantly reduced in roots and leaves of both transgenic P+T-NL and P+T-KY lines. **A-C:** reduction of nicotine (**A**), nornicotine (**B**), and total nicotine and nornicotine (**C**) in leaves of three P+T-NL lines (1, 2, and 3); **D-F:** reduction of nicotine (**D**), nornicotine (**E**), and total nicotine and nornicotine (**F**) in roots of three P+T-NL lines (1, 2, and 3); **G-I:** reduction of nicotine (**G**), nornicotine (**H**), and total nicotine and nornicotine (**I**) in leaves of three P+T-KY lines (1, 2 and 3); **J-L:** reduction of nicotine (**J**), nornicotine (**K**), and total nicotine and nornicotine (**L**) in

roots of three P+T-NL lines (1, 2, and 3). Bars labeled with “a” and “b” means p-value less than 0.05 and bars labelled with the same low case “a” or “b” means no significant difference.

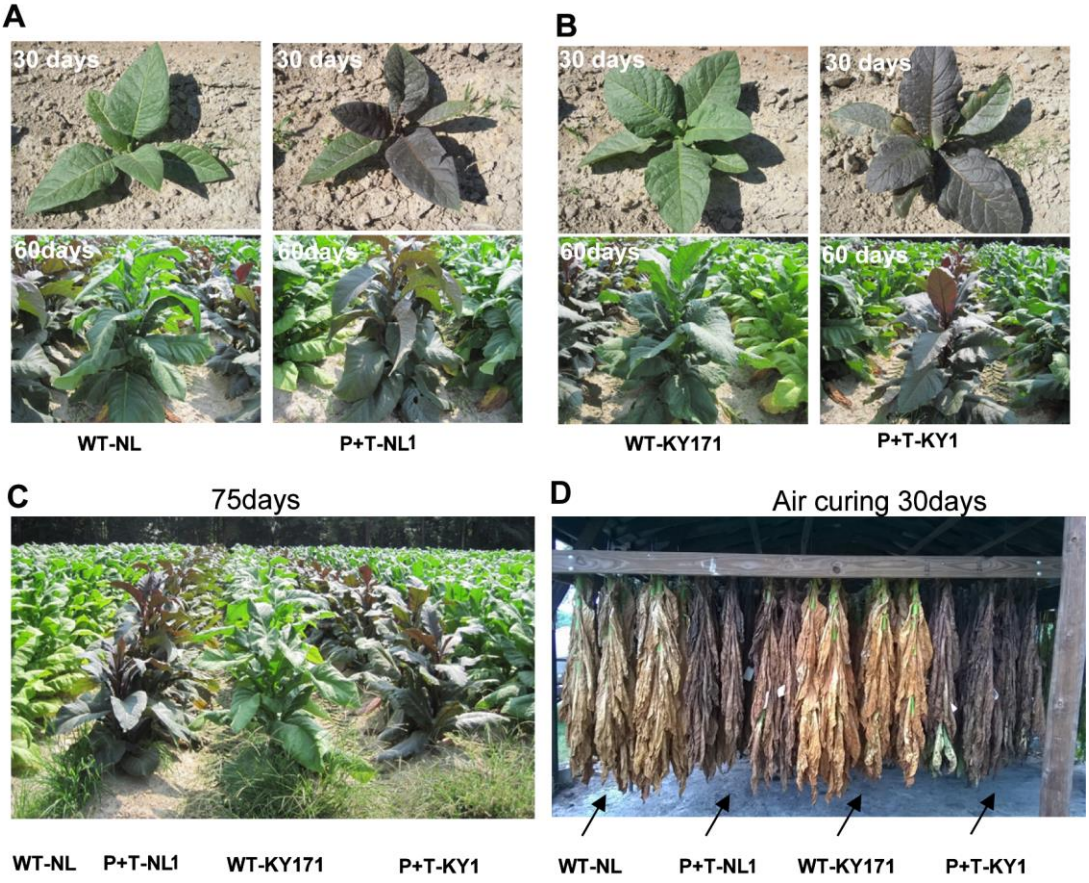


Figure 5 Phenotypes of P+T-NL1 and P+T-KY1 plants versus their corresponding wild-type tobacco plants in the field and of air-cured leaves. Field farming practice of four genotypes was performed in Oxford, North Carolina. **A**, phenotypes of 30-day and 60-day old WT-NL vs. red P+T-NL1 plants. **B**, phenotypes of 30-day and 60-day old WT-KY171 vs. red P+T-KY1 plants. **C-D**, phenotypes of topped plants (**C**) and leaves in air curing (**D**). Plant name abbreviations are WT-NL: wild-type Narrow Leaf Madole (NL) variety, P+T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WT-KY171: wild-type KY171 variety, and P+T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.

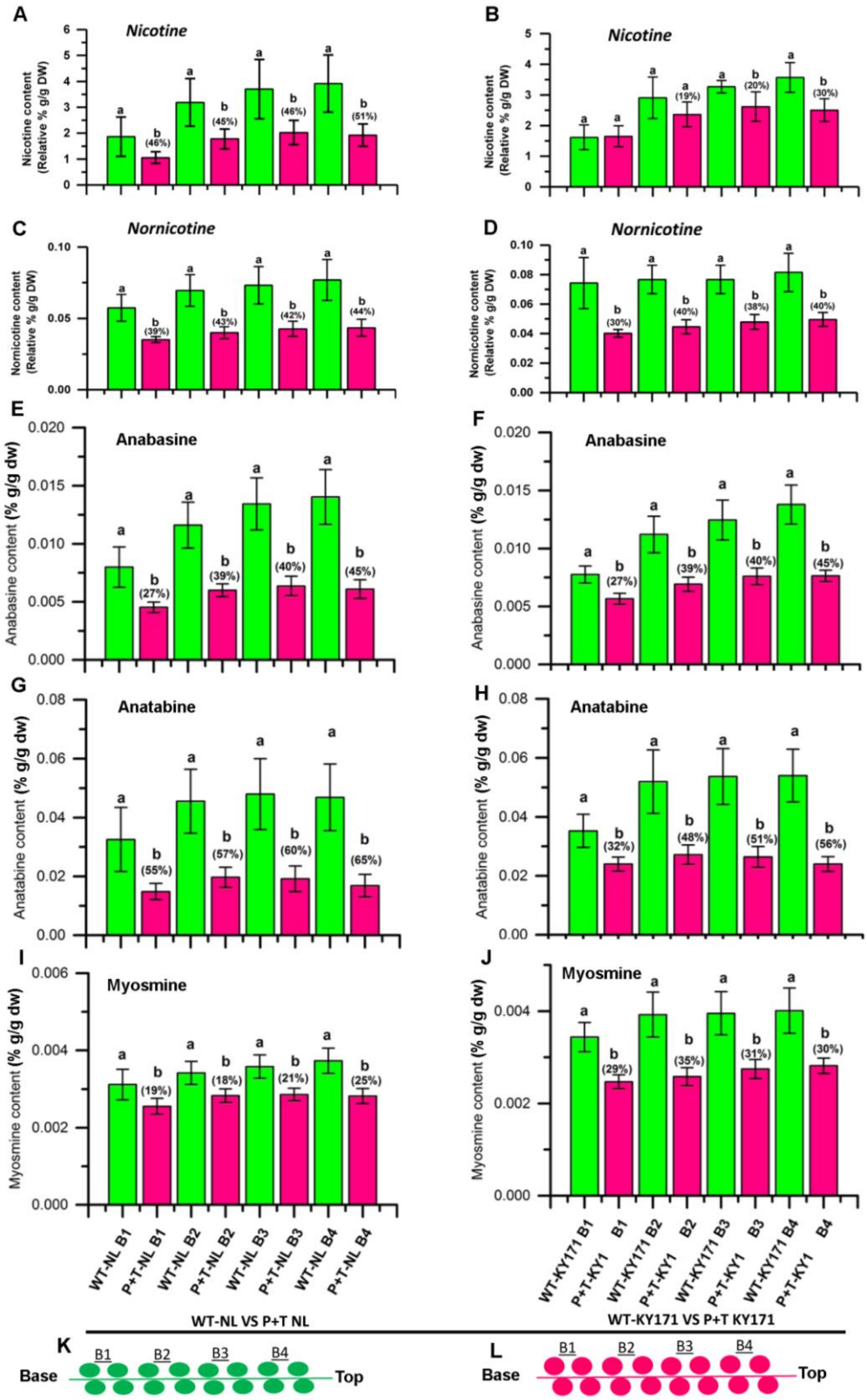
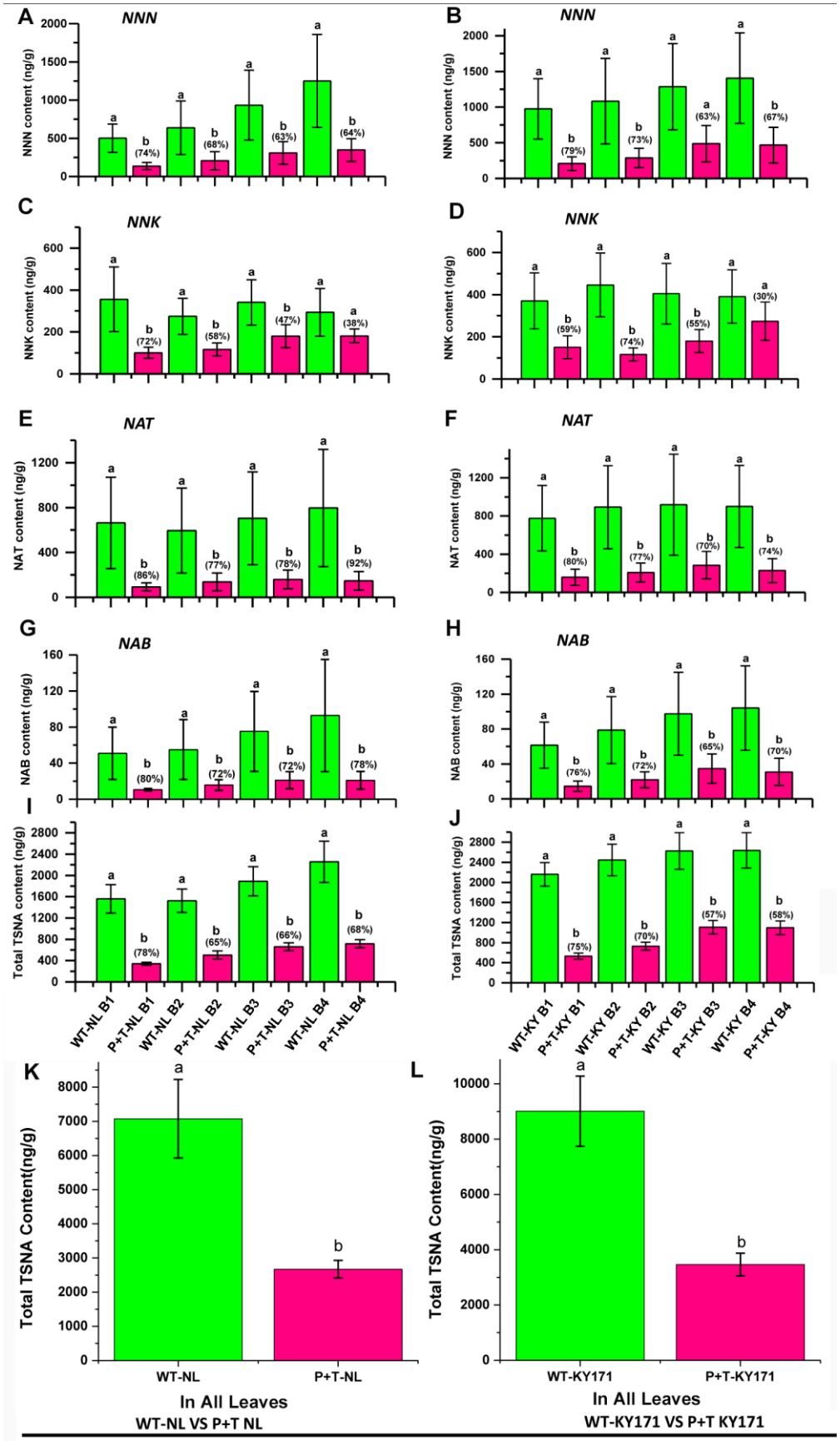


Figure 6 Reduction of nicotine and four other alkaloids in cured leaves of red P+T-NL1 and P+T-KY1 tobacco plants. A-J, reduction of nicotine (**A-B**),

nornicotine (**C-D**), anabasine (**E-F**), anatabine (**G-H**), and myosmine (**I-J**) in four leaf groups of P+T-NL (**A, C, E, G, and I**) and P+T-KY plants (**B, D, F, H, and J**). **K-L**, cartoons showing leaf groups harvested from both wild type (**K**) and transgenic red plants (**L**).

Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of tobacco plant, respectively. Each metabolite in each leaf group was quantified from leaves harvested from 120 tobacco plants in three plots. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Bars labelled with low case “a” and “b” mean significant difference between paired green and red bars (n=120, P-value less than 0.05). Percentages labeled in the top of red bars mean reduction compared to green bars. The standard bars mean standard deviation. Abbreviations, WT-NL: wild-type Narrow Leaf Madole (NL) variety, P+T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WT-KY: wild-type KY171 variety, and P+T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.


A
NNN
NNN content (ng/g)
B
NNN
NNN content (ng/g)
C
NNK
NNK content (ng/g)
D
NNK
NNK content (ng/g)
E
NAT
NAT content (ng/g)
F
NAT
NAT content (ng/g)
G
NAB
NAB content (ng/g)
H
NAB
NAB content (ng/g)
I
Total TSNA content (ng/g)
J
Total TSNA content (ng/g)
WT-NL B1
P+T-NL B1
WT-NL B2
P+T-NL B2
WT-NL B3
P+T-NL B3
WT-NL B4
P+T-NL B4
WT-KY B1
P+T-KY B1
WT-KY B2
P+T-KY B2
WT-KY B3
P+T-KY B3
WT-KY B4
P+T-KY B4
K
Total TSNA Content(ng/g)
WT-NL
P+T-NL
In All Leaves
WT-NL VS P+T NL
L
Total TSNA Content(ng/g)
WT-KY171
P+T-KY171
In All Leaves
WT-KY171 VS P+T KY171

1694

Figure 7 Reduction of four individual and total TSNAs in leaves of red P+T-NL1 and P+T-KY1 tobacco plants. A-H, reduction of NNN (**A-B**), NNK (**C-D**), NAT (**E-F**), and NAB (**G-H**) in each leaf group of P+T-NL1 (**A, C, E,** and **G**) and P+T-KY1 (**B, D, F,** and **H**) plants; **I-J,** reduction of total TSNA contents summed from four individual TSNAs of each leaf group of P+T-NL1 (**I**) and P+T-KY1 (**J**) plants; **K-L,** reduction of total TSNA contents summed from four individual TSNAs of all leaves of P+T-NL1 (**K**) and P+T-KY1 (**L**) plants.

Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of plants, respectively. Each metabolite in each group of leaves was quantified from 120 tobacco plants. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Paired green and red bars labelled with low case “a” and “b” mean significant difference (n=120, P-value less than 0.05). Percentages labeled on the top of red bars indicate reduction levels compared to green bars. Abbreviation, WT-NL: Narrow Leaf Madole, WT-KY171: wild type KY171, P+T-NL: stacked PAP1 and TT8 transgenic NL line 1, and P+T-KY1: stacked PAP1 and TT8 transgenic KY171 line 1.

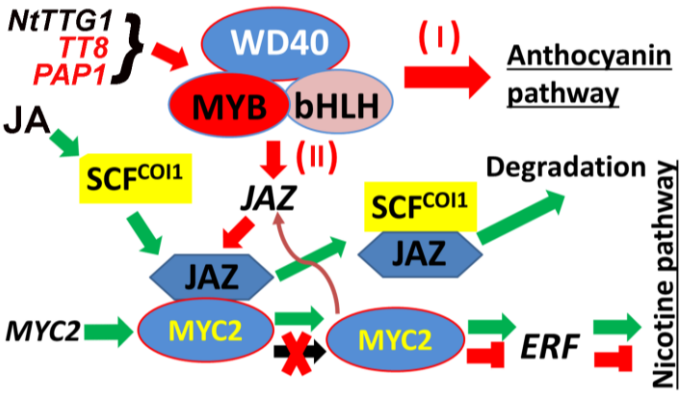


Figure 8 A model showing the distant pathway-cross regulation (DPCR) created for the downregulation of the biosynthesis of tobacco alkaloids in red tobacco plants

Regulation pathways created in red tobacco plants are featured by red colored solid arrows, "X" shape, and "T" blocks. The stacked *PAP1* and *TT8* as well as the endogenous *NtTTG1* encode MYB, bHLH, and WD40 proteins, respectively, which form a complex to regulate two biosynthetic pathways. The regulation pathway (I) is the activation of biosynthesis of anthocyanin. The regulation pathway (II) is the DPCR that the MYB-bHLH-WD40 complex upregulates the endogenous *NtJAZs* leading to the reduction of MYC2, downregulation of *EAR*, and the nicotine biosynthesis. In addition, the JA-JAZ-MYC regulatory pathway is indicated with green arrows starting with JA (greenish arrows) and the feedback loop (indicated by a brownish bending arrow). SCF: Skp/Cullin/F-box and COI1: coronatine insensitive 1.