Allelic diversity of NAC18.1 is a major determinant of fruit firmness and harvest date in apple (Malus domestica)

Trevor H. Yeats¹,⁶,⁷, Zoë Migicovsky², Sophie Watts², Jun Song³, Charles F. Forney³, Karen Burgher-MacLellan³, Daryl J. Somers⁴, Julia Vrebalov⁵, James G. Giovannoni⁵,⁶, Jocelyn K. C. Rose¹*, Sean Myles²*

¹. Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853 USA
². Department of Plant, Food and Environmental Sciences, Dalhousie University, Faculty of Agriculture, Agricultural Campus, Truro, NS, Canada.
³. Agriculture and Agri-Food Canada, AFHRC, Kentville, Nova Scotia, B4N 1J5, Canada
⁴. Vineland Research and Innovation Centre, Vineland Station, ON, L0R 2E0, Canada
⁵. Boyce Thompson Institute, Cornell University, Ithaca, NY 14853 USA.
⁶. United States Department of Agriculture, Robert W. Holley Center, Cornell University, Ithaca, NY 14853 USA
⁷. Present address: Plant Breeding & Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853 USA

* Co-corresponding authors: Sean Myles (seanmichaelmyles@dal.ca) and Jocelyn K. C. Rose (jr286@cornell.edu).
Highlight:

*NAC18.1* is a member of a family of conserved transcriptional regulators of ripening that underlies variation in fruit firmness and harvest date in diverse apple accessions.

Abstract:

Softening is a hallmark of ripening in fleshy fruits, and has both desirable and undesirable implications for texture and postharvest stability. Accordingly, the timing and extent of ripening and associated textural changes are key targets for improving fruit quality through breeding. Previously, we identified a large effect locus associated with harvest date and firmness in apple (*Malus domestica*) using genome-wide association studies (GWAS). Here, we present additional evidence that polymorphisms in or around a transcription factor gene, *NAC18.1*, cause variation in these traits. First, we confirmed our previous findings with new phenotype and genotype data from ~800 apple accessions. In this population, we compared *NAC18.1* to three other ripening-related markers currently used by breeders (*ACS1, ACO1*, and *PGI*), and found that the effect of the *NAC18.1* genotype on both traits greatly exceeded that observed for the other markers. By sequencing *NAC18.1* across 18 accessions, we revealed two predominant haplotypes containing the SNP previously identified using GWAS, as well as dozens of additional SNPs and indels in both the coding and promoter sequences. *NAC18.1* encodes a protein with high similarity to the NON-RIPENING (NOR) transcription factor, an early regulator of ripening in tomato (*Solanum lycopersicum*). To test whether these genes are functionally orthologous, we introduced *NAC18.1* transgenes into the tomato *nor* mutant and showed that both haplotypes complement the *nor* ripening deficiency. Taken together, these
results indicate that polymorphisms in NAC18.1 underlie substantial variation in apple firmness and harvest time through modulation of a conserved ripening program.

**Keywords:**

Apple, fruit ripening, NAC-domain transcription factor

**Abbreviations:**

| Abbreviation | Description                                      |
|--------------|--------------------------------------------------|
| ABC          | Apple Biodiversity Collection                    |
| ACS1         | 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 1     |
| ACS2         | 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 2     |
| ACO1         | AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 1        |
| CIELAB       | International Commission on Illumination 1976 color space |
| HRM          | High Resolution Melting                          |
| LD           | Linkage disequilibrium                           |
| PG1          | POLYGALACTURONASE 1                              |
| PG2          | POLYGALACTURONASE 2                              |
| PSY1         | PHYTOENE SYNTHASE 1                              |
| REML         | Restricted Maximum Likelihood                    |
| RPL2         | RIBOSOMAL PROTEIN L2                             |
| USDA         | United States Department of Agriculture          |
Despite their diverse structure, ontogeny, and biochemical composition, fleshy fruits from a taxonomically broad range of species undergo coordinated ripening programs that have many features in common. Ripening involves numerous physiological and biochemical changes that render the fruit attractive and nutritious for consumption by seed-dispersing animals, or human consumers in the case of cultivated crops. These include the ripening-associated accumulation of sugars, pigments, and flavor or aroma compounds, as well as a loss of flesh firmness due in large part to the controlled modification and depolymerization of cell wall polysaccharides (Wang et al., 2018). These processes are regulated by conserved and convergently evolved networks of transcription factors and hormones, such as ethylene in so-called climacteric fruit (Lü et al., 2018).

Various aspects of fruit ripening have been particularly well studied in tomato (*Solanum lycopersicum*) and the characterization of tomato ripening mutants has revealed a regulatory network, consisting of transcription factors, hormones, and epigenetic modifications (Giovannoni et al., 2017). Of the genes that regulate ripening, the *NON-RIPENING (NOR)* gene appears to exhibit the earliest expression preceding ripening (Shinozaki et al., 2018). *NOR* encodes a NAC domain transcription factor (Giovannoni et al., 2004). This family of transcription factors is one of the largest plant-specific families of transcription factors, with specific members regulating development, defense, and senescence (Mathew and Agarwal, 2018). While all members share a conserved DNA-binding (NAC) domain, specific functional clades are defined in terms of their more variable domains, particularly the C-terminal transcriptional regulatory region. These domains can act directly as transcriptional activators, or can facilitate interaction with other transcription factors in order to fine-tune transcriptional control.
Apple (*Malus domestica*) fruit exhibit extensive variation in the extent and timing of ripening and ripening-associated softening. This is reflected in three related agronomic traits: harvest date, firmness, and storability. Harvest date is typically determined by the semi-quantitative measurement of starch, based on iodine-staining patterns (Blanpied and Silsby, 1992). Firmness can also be used as a measure of ripening, although apples soften less than many other fleshy fruits, and are more noted for their characteristic and desirable “crisp” texture. Importantly, postharvest storability is highly correlated with fruit firmness (McClure *et al.*, 2018; Ornelas-Paz *et al.*, 2018). Thus, firmness is typically a desirable fruit quality trait in apple, although it is often associated with later harvest dates that can be undesirable if the growing season is short (Migicovsky *et al.*, 2016).

Due to a prolonged juvenile phase, it is particularly challenging to evaluate fruit quality traits in the context of an apple breeding program. As a result, in recent years considerable effort has been invested in developing molecular markers that can be used to select for fruit quality traits at the seedling stage. In particular, three markers have been proposed for predicting firmness. The first is *1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 1 (ACS1)*, which encodes the ripening-associated isoform of an ethylene biosynthesis gene. The *ACS1*-2 allele contains a retrotransposon insertion thought to confer low ethylene production and longer shelf life in Fuji and other apple cultivars homozygous for this allele (Sunako *et al.*, 1999; Harada *et al.*, 2000). The second is a marker for a gene corresponding to another enzyme involved in ethylene biosynthesis, *AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 1 (ACO1)*; Costa *et al.*, 2005). Finally, a marker has been developed for alleles of *POLYGALACTURONASE 1 (PG1)*; Costa *et al.*, 2010). This gene encodes an enzyme that
hydrolyzes pectin polysaccharides and has been implicated in apple fruit firmness by RNAi-based gene silencing experiments (Atkinson et al., 2012).

In a previous study, we conducted GWAS on a range of traits in apple using genotype data and historical phenotype records for 689 accessions of the United States Department of Agriculture (USDA) apple germplasm collection (Migicovsky et al., 2016). Our analysis revealed a single nucleotide polymorphism (SNP) on chromosome 3 within the coding sequence of the NAC18.1 gene that was significantly associated with firmness and harvest date. This SNP results in an aspartate (D) to tyrosine (Y) mutation at position 5 of the NAC18.1 amino acid sequence and we refer to this SNP as D5Y (Migicovsky et al., 2016). Subsequently, GWAS of two European apple germplasm collections confirmed the association between NAC18.1 and ripening time (Urrestarazu et al., 2017; Larsen et al., 2019). NAC transcription factors have also been linked to harvest date in other Rosaceae species, including peach (Prunus persica; Pirona et al., 2013) and apricot (P. armeniaca; García-Gómez et al., 2019).

Here we report an extension of our previous findings using three approaches. First, we conducted additional phenotyping and genotyping of apple germplasm in order to compare the predictive power of the NAC18.1 D5Y SNP with other proposed firmness/harvest date markers. Secondly, we sequenced the NAC18.1 gene from a subset of apple cultivars in order to discover additional polymorphisms in linkage disequilibrium (LD) with D5Y. Finally, we used heterologous expression of the NAC18.1 gene in the tomato non-ripening (nor) mutant to test whether it functions as a component of a conserved fruit ripening program.
Materials and Methods

Apple Firmness and Harvest Date

The Apple Biodiversity Collection (ABC) is an apple orchard located in Kentville, Nova Scotia, which contains 1,113 accessions. It was first established in spring 2012, when trees were budded onto M.9 rootstocks. In fall 2012, the trees were uprooted and kept in cold storage, before planting in spring 2013. The trees were planted in an incomplete block design, which includes 1 of 3 standards per grid, allowing for correction of positional effects using a REstricted Maximum Likelihood (REML) model, described in Migicovsky et al. (2018).

In 2017, we evaluated harvest date for 1,348 trees and fruit firmness for 1,328 trees within the ABC orchard. Due to the diversity of apples within the collection, a variety of methods were used to determine the appropriate time to harvest. First, we observed if the tree had dropped fruit or, for red apples, if the fruit were a deep red color. Next, a sample apple was taken from each tree and touched to assess firmness, tasted to assess starch and sweetness, cut in half to check browning of seeds, and then sprayed with iodine solution to assess starch content (Blanpied and Silsby, 1992).

When fruit were determined to be mature, they were harvested and evaluated for firmness. We recorded the firmness (kg cm$^{-2}$) of 5 fruit per tree using a penetrometer with a 1 cm diameter (Fruit Texture Analyzer, GS-14, Güss Manufacturing). A small section of skin was removed using a vegetable peeler, and each fruit was placed on the penetrometer platform so that the piston entered the middle of the apple where the skin had been removed. Data were automatically recorded into a spreadsheet.

Due to the number of trees, harvesting the ABC orchard often lasted more than one day. Therefore, differences in harvest date within a week were simply due to the time required to
harvest, and the harvest date for each tree was recorded as the Monday of that week. We used the “lmer” function in the R package lme4 (Bates et al., 2015) to fit a REML model for harvest and firmness data. Next, we calculated the least squares mean using the “lsmeans” function in the lsmeans R package (Lenth, 2016), resulting in one value per unique accession. We fit this model for 866 unique accessions with harvest dates and 863 accessions with firmness measurements.

**High Resolution Melting (HRM) Genotyping**

DNA was extracted using silica columns from fresh leaf tissue collected from the ABC orchard, quantified using PicoGreen (Thermo) and normalized to a concentration of 20 ng µL⁻¹. Genotyping was conducted using PCR and high resolution melting (HRM) on a LightScanner HR384 (BioFire). Primers are listed in Supplemental Table S1.

Genotyping of the PG1 SNP marker was based on the GenBank sequence L27743.1, where the G allele at position 437 is unfavorable and the T allele is favorable (Costa et al., 2010). Three allelic combinations of the observed indel, ACS1-1/1, 1-1/2 and 1-2/2, have been associated with high, medium, and low ethylene production, respectively (Sunako et al., 1999; Harada et al., 2000; Oraguzie et al., 2004; Costa et al., 2005). The exact position of the indel is from ACS1-2 (GenBank: AB010102.1) 1,320 to 1,483 bp (163 bp), and from ACS1-1 (GenBank: AY062129.1) 4,500 to 4,525 bp (25 bp). The size difference between alleles is 138 bp. The ACO marker involves an unfavorable 62 bp insertion in the third intron of ACO1 (Costa et al., 2005). The third intron spans from 1,083 bp to 1,300 bp of the GenBank sequence Y14005.1 and the indel is found from position 1,297 bp to position 1,358 bp. The D5Y mutation in NAC18.1 is a nonsynonymous SNP at position 30,698,039 on chromosome 3, according to reference genome version GDDH13 v1.1 (https://iris.angers.inra.fr/gddh13/) and is associated
with both harvest date and firmness (Migicovsky et al., 2016). The desirable C allele encodes an aspartic acid (D) at the fifth amino acid position of the NAC18.1 protein, while the undesirable A allele encodes a tyrosine (Y).

Marker-phenotype associations

We evaluated the ability of markers at NAC18.1, ACO1, ACS1, and PG1 to predict both harvest date and firmness of accessions in the ABC orchard. The number of accessions with phenotype and genotype information ranged from 754 to 852 depending on the trait/marker combination. The association test was conducted using Spearman’s rank correlation test. We visualized results using the “geom_boxplot” function in ggplot2 in R (Wickham, 2016). Lastly, we tested which mode of inheritance best fit the data using SNPstats (Solé et al., 2006), and selected the model with the lowest Akaike Information Criterion value.

Sequencing of NAC18.1

DNA was isolated from leaves sampled from 18 apple accessions growing in Geneva, NY. A 2.3 kb amplicon including the NAC18.1 gene and ~800 bp of upstream sequence was amplified by PCR using primers NAC18F2 and NAC18R2 (Supplemental Table S1) and Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB). PCR product size and purity was confirmed by agarose gel electrophoresis, and the remaining product was purified using DNA Clean & Concentrator kit (Zymo Research). The resulting DNA fragment was cloned into the plasmid pMiniT 2.0 and transformed into E. coli using the NEB® PCR Cloning Kit (NEB).

Individual colonies were selected for complete sequencing of the cloned amplicon using the primers NAC18F2, NAC18F3, NAC18F4, NAC18R1, and NAC18R2 (Supplemental Table
For accessions homozygous for the D5Y SNP, the NAC18.1 amplicon from a single clone was sequenced. For heterozygous accessions, two clones representing each D5Y allele were selected based on partial sequencing of the D5Y region, followed by complete sequencing of the 2.3 kb amplicon, as described above. The nucleotide sequences were aligned by MUSCLE (Edgar, 2004) and used to construct a maximum-likelihood phylogenetic tree in MEGA7 (Kumar et al., 2016).

**Molecular cloning and plant transformation**

The hypothetical coding sequences (CDSs) corresponding to the consensus sequence of the “A” and “C” haplotypes of NAC18.1 were synthesized as a double-stranded DNA (gBlock) by IDT, with 20 bp of flanking sequence added to facilitate Gibson Assembly of the NAC18.1 sequences between the AscI and PacI restriction sites of pMDC32 (Curtis and Grossniklaus, 2003). Constructs were assembled using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB) and their integrity verified by Sanger sequencing. Plasmids were transformed into Agrobacterium tumefaciens, which was used to transform tomato cotyledon explants (Van Eck et al., 2019) derived from the tomato nor mutant in the cv. Ailsa Craig background (LA3770, Tomato Genetics Resource Center, https://tgrc.ucdavis.edu/). A total of 10 T0 plants were recovered per construct and plants derived from two independent transformation events per construct were selected for further characterization in the T1 generation.

**Transgenic tomato characterization**

Fruit were harvested when visually ripe, or in the case of the nor mutant, at the equivalent age as ripe cv. Ailsa Craig fruit (the near isogenic wild-type control), as determined by tagging
of flowers at anthesis. The color of the fruit surface was measured using a CR-400 Chroma Meter (Konica Minolta), and fruit were weighed, photographed, and dissected. Pericarp tissue was frozen in liquid nitrogen and stored at -80°C. Frozen tissue was ground to a fine powder and RNA extracted by a modified version of Chang et al. (1993). Briefly, approximately 400 mg of tissue was added to a preheated (80°C) two-phase system consisting of 500 µL of water-saturated phenol and 500 µL of extraction buffer (100 mM Tris [pH 8.0], 25 mM EDTA, 2 M NaCl, 2% CTAB, 2% PVP, and 2% [v/v] beta-mercaptoethanol). The mixture was vortexed and incubated for 5 min at 65°C and then cooled to room temperature before extracting and precipitating RNA, as previously described (Chang et al., 1993).

RNA was treated with RNase-free DNase (Promega) and used for cDNA synthesis with RNA to cDNA EcoDry™ Premix with Oligo dT primer (Takara Bio). The cDNA was used as template for quantitative PCR using Luna Universal qPCR Master Mix (NEB) and a Viia7 real-time PCR instrument (Life Technologies/ABI). Gene-specific primers are listed in Supplemental Table S1. Quantification used the ∆Ct method with RIBOSOMAL PROTEIN L2 (RPL2) as a reference gene, and statistical significance of the ∆Ct values was tested by a one-way ANOVA followed by Tukey’s HSD test.

Results

Evaluation of markers for firmness and harvest date in apple

We sought to re-evaluate the ability of D5Y and other published markers to predict fruit firmness and harvest date using new phenotype and genotype data: specifically, we aimed to address some experimental limitations of our previous work. First, the GWAS that identified D5Y, and failed to find associations for previously published markers, made use of only 8,000
SNPs (Migicovsky et al., 2016). Given the rapid LD decay in apple, our low SNP density was insufficient to conclude that we had exhaustively searched the apple genome for loci involved in firmness and harvest date (Migicovsky et al., 2016). Secondly, the historic phenotype data we used were imprecise. For example, firmness was recorded as either “firm” or “soft” rather than as a biomechanical measurement, and harvest date was recorded as simply “early”, “mid”, or “late”. These factors would have limited the detection of additional loci that underlie firmness and/or harvest date.

In order to directly compare the NAC18.1 D5Y marker that we identified with previously published markers, we developed HRM genotyping assays for the D5Y SNP, as well as markers at ACS1, ACO1, and PG1. We used these assays to genotype accessions from the ABC orchard (Supplementary Table S2). While this orchard largely consists of clones of individuals present in the population used in our previous GWAS (i.e. the USDA germplasm collection), it also contains a number of additional accessions. More importantly, for the purposes of phenotyping, the ABC orchard is planted in an incomplete block design that allows for modeling of location effects using REML methods (Migicovsky et al., 2018). Using this approach, we generated firmness and harvest date measures for over 800 unique accessions for which HRM genotype data were available.

We evaluated the mode of inheritance for each genetic marker and phenotype combination and found that, while the effects of the markers at ACS1, ACO1, and PG1 were all dominant, the D5Y marker at NAC18.1 had a codominant effect. As a result, accessions that were heterozygous for D5Y had firmer fruit (an increase of 1.2 kg cm$^{-2}$) and a later harvest date (20.13 days) than those that were homozygous for the ‘soft’ A allele. The combined effect of two C alleles was notably stronger than the effect of a single C allele, increasing firmness by 2.24 kg
cm² and harvest date by a month (30.52 days), when compared to AA genotypes. The difference between firm/late and soft/early genotypes for D5Y was at least 4 times higher for firmness and 3 times higher for harvest date than for the markers at *ACS1, ACO1*, and *PG1* (Figure 1).

Figure 1. Distribution of phenotypic variation across genotypic classes for markers at *NAC18.1, PG1, ACO1*, and *ACS1*. A. Boxplots of variation in firmness across each marker B. Boxplots of variation in harvest date across each marker. The lower and upper hinges of the
boxplots correspond to the first and third quartiles. The putative inheritance model, allele effect, correlation (rho) and \( p \)-value determined using a Spearman’s rank correlation test, and the number of accessions within each genotypic class are listed.

To evaluate the distribution of firmness/harvest date markers across commercial cultivars, we present the genotypes for all markers genotyped here in nine of the top ten apple cultivars sold in the USA (Table 1). All nine cultivars were homozygous for the favorable (firm) C allele of \( NAC18.1 \), while only 2 to 4 of the cultivars had homozygous favorable genotypes for the other markers. Among the top cultivars, only ‘Fuji’, released in 1962, was homozygous for favorable alleles at all markers, while ‘McIntosh’, initially discovered in 1811 and commercially released in 1870, was homozygous for undesirable alleles at all markers except for the \( NAC18.1 \) marker.
Table 1. Genotypes of markers for 9 of the 10 most widely sold apple cultivars in the United States.

| Release Date | Cultivar     | Genotypes<sup>a</sup> |
|--------------|--------------|------------------------|
|              |              | **NAC18.1** | **ACS1** | **ACO1** | **PG1** |
| 1868         | Granny Smith | +           | ±        | –        | +       |
| 1870         | McIntosh     | +           | –        | –        | –       |
| 1895         | Red Delicious| +           | ±        | +        | ±       |
| 1914         | Golden Delicious | +   | ±        | –        | ±       |
| 1952         | Braeburn     | +           | ±        | n.d.     | ±       |
| 1962         | Fuji         | +           | +        | +        | +       |
| 1974         | Gala         | +           | +        | ±        | ±       |
| 1990         | Cripps Pink  | +           | ±        | –        | +       |
| 1991         | Honeycrisp   | +           | ±        | –        | +       |

<sup>a</sup> Genotypes are coded as + for homozygous, favorable; – for homozygous, unfavorable; ± for heterozygous.

Resequencing of **NAC18.1**

Since the D5Y SNP is predicted to result in an amino acid change in the NAC18.1 protein and LD decay is generally high in apple (r<sup>2</sup> decayed to < 0.2 within 100 bp), it is possible that D5Y is causative for the early/firm phenotype (Migicovsky et al., 2016). However, GBS sequence data are too sparse to rule out additional polymorphisms in LD with D5Y contributing to the phenotype. To better understand allelic diversity of **NAC18.1**, we selected 6 cultivars each of the D5Y A/A, A/C, and C/C genotypes, based on our HRM genotyping data. For homozygous...
cultivars, single alleles were sequenced, while for the heterozygous samples we resequenced both alleles, resulting in NAC18.1 sequences from 24 haplotypes. These were compared to two reference sequences from the GDDH13 v1.1 apple genome (Daccord et al., 2017): Md03g1222600 and Md11g1239900, representing NAC18.1 and its closest paralog, respectively.

In addition to confirming the expected D5Y genotype in all individuals, a number of additional SNPs and indels were revealed within both coding and non-coding regions of the NAC18.1 allele. Multiple sequence alignment and subsequent phylogenetic analysis indicated two major clades corresponding to the D5Y A and C genotypes (Figure 2A). SNPs and small indels were observed throughout the sequenced region. Notably, in addition to the D5Y amino acid change, several additional amino acid changes were observed between sequences from D5Y genotypes and the reference sequence of NAC18.1. For example, near the site of the D5Y polymorphism, all “A” haplotypes also had a 12 nucleotide insertion that introduced the amino acid sequence QPQP (Figure 2B).

Taken together, these results indicate that the NAC18.1 locus exhibits a number of polymorphisms that are in strong LD with each other across the 2.3 kb that we sequenced. Many of these are candidates for causality, as they affect either the coding sequence or promoter of the gene. On the other hand, it is possible the signal at D5Y is driven by a causal variant outside of the 2.3kb region sequenced here that acts completely independently from NAC18.1. This led us to look for further evidence of the involvement of NAC18.1 in the ripening process in apple.
Figure 2. Polymorphisms in LD with the D5Y SNP. A. Maximum likelihood phylogenetic tree of 24 NAC18.1 sequences from 18 apple cultivars and the reference genome (GDDH13 v1.1) sequences of NAC18.1 (Md03g1222600) and its closest homolog (Md11g1239900). A graphical representation of the 2.3 kb multiple sequence alignment is also shown. B. Amino acid sequence alignment of the N-terminal region of NAC18.1, illustrating additional variation in the coding sequence strongly correlated with the D5Y variant. The complete amino acid sequence alignment is given in Supplemental Figure S1.

Transgenic complementation of the tomato non-ripening (nor) mutant

A BLAST search of predicted proteins from the ITAG 3.20 tomato genome (https://solgenomics.net/) using NAC18.1 as a query identified only two sequences matching the full length of the query: Solyc10g006880 (NOR) and Solyc07g063420 (NOR-LIKE 1). To test the functional conservation of NAC18.1 and NOR, we introduced transgenic constructs individually conferring constitutive expression of each of the NAC18.1 haplotype CDSs into the tomato nor
mutant. Two independent lines for each construct were characterized in the T1 generation with respect to their ability to rescue the ripening deficiency of the nor mutant. In contrast to nor, all four lines changed color at maturity, although internal color change did not occur to the same extent as observed in a WT control (Figure 3A). To complement this qualitative phenotypic assessment, we also conducted quantitative colorimetry of the surface of fruits (Figure 3B). Fruit from all the transgenic lines exhibited a significant increase in the a* (green-red) component of CIELAB (International Commission on Illumination 1976) color space relative to the nor mutant, although only NAC18.1 A #6 achieved similar a* levels to WT.

The different degrees to which ripe fruit color was restored in the transgenic plants might be a consequence of either the different alleles of NAC18.1, or of different levels of transgene expression in each line. To address this, we analyzed the expression level of NAC18.1 using qRT-PCR primers designed to target both NAC18.1 alleles. Expression levels of the NAC18.1 transgene were not statistically different (p > 0.5) between each independent line (Figure 4A). Next, we measured the expression level of several genes associated with tomato ripening physiology: PHYTOENE SYNTHASE 1 (PSY1) encodes an early enzyme of carotenoid synthesis responsible for the production of red pigments during fruit ripening (Fray and Grierson, 1993). Consistent with the green color of mature nor fruit, expression of PSY1 is impaired in the nor mutant (Osorio et al., 2011); POLYGALACTURONASE 2 (PG2) encodes an enzyme catalyzing pectin depolymerization associated with fruit softening (Biggs and Handa, 1989); and 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 2 (ACS2) encodes an enzyme that synthesizes 1-aminocyclopropane-1-carboxylate, the immediate precursor of ethylene (Nakatsuka et al., 1998).
Expression of all three marker genes was enhanced in the NAC18.1 transgenic lines relative to the nor mutant control, although not to the same extent as observed in WT ripe fruit (Figure 3B-D). In contrast to the consistent level of NAC18 expression observed in each line, the marker genes were more variable in their expression level between lines. A similar pattern was observed for all marker genes, with the NAC18.1C #9 line showing the smallest induction of marker gene expression relative to nor. In the case of PG2 and ACS2, the difference in expression in NAC18.1C #9 was not statistically significant relative to nor (p = 0.09 and 0.18, respectively). Consistent with these results, fruit from this line also exhibited the lowest amount of color development (Figure 3B). Taken together, these results indicate that a canonical ripening program can be induced in the tomato nor mutant through the heterologous expression of either apple NAC18.1 allele.
Figure 3. Transgenic complementation of the tomato nor mutant. A. Mature fruits of the tomato nor mutant and four independent T1 transgenic lines constitutively expressing either of two alleles of the NAC18.1 transgene, NAC18.1C and NAC18.1A and isogenic WT control (cv.
Ailsa Craig). B. Quantitative colorimetry of the fruit surface of nor, the transgenic fruit and WT fruit. The a* component (green-red axis) is shown, with the mean ± SE superimposed in black over the raw values in gray (n = 5). Genotypes not sharing a letter (a-e) are statistically distinct by one-way ANOVA and Tukey’s HSD test (p < 0.05).

**Discussion**

Genomics-assisted breeding has tremendous potential in perennial crops like apple, where a lengthy juvenile phase and large plant size make phenotyping at the adult stage time-consuming and expensive. However, the genetic markers used for culling progeny at the seedling stage must accurately predict the trait of interest at the adult stage in order for genomics-assisted breeding to be effective (Luby and Shaw, 2001). Firmness and harvest date, which are highly correlated, are key breeding targets in apple, and markers in three genes (ACS1, ACO1, and PG1) are widely used in genomics-assisted apple breeding programs to predict firmness. However, recent GWAS analyses for firmness and harvest date have failed to identify significant associations at any of these loci, but rather have uncovered a single major effect locus at the NAC18.1 gene (Migicovsky et al. 2016; McClure et al. 2018; Larsen et al. 2019). Based on the results of this current study, we propose several possible explanations for these findings.

First, the markers in ACS1, ACO1, and PG1 were identified using linkage mapping in biparental crosses, which only captured variation segregating in the progeny of those crosses. These mapping studies generally made use of commercially successful cultivars, and here we show that the most commercially successful apple cultivars in the USA are all homozygous for the desirable NAC18.1 alleles (Table 1). Thus, the NAC18.1 alleles we identified via GWAS may
not have segregated in the progeny of the crosses used to discover the markers at *ACS1*, *ACO1* and *PG1*. This may be the primary reason *NAC18.1* was not identified using traditional linkage mapping approaches. Second, the most extensive GWAS we have performed to date may have failed to detect association signals at *ACS1*, *ACO1* and *PG1* since the SNP density used (~1 SNP every 100kb) provided poor genome-wide coverage, given the rapid LD decay observed in our GWAS population ($r^2$ decayed to < 0.2 within 100 bp). For example, the SNP closest to *ACO1* in our GWAS was >100 kb away and was thus unlikely to be in LD with genetic variants within *ACO1* (Migicovsky et al. 2016). Given these limitations, we examined the value of each of these markers to predict firmness and harvest date in a diverse collection of over 750 unique apple accessions.

The correlation test between each of the markers and the two phenotypes, harvest date and firmness, revealed statistically significant ($P < 0.05$) associations in every case except one (Figure 1). ACO1 was not correlated with firmness in our data, which is consistent with a recent study that found that a marker for ACO1 did not explain any of the variation in crispness or firmness across 207 accessions from New Zealand’s apple breeding program (Chagne et al., 2019). Our observation that most of the marker-trait tests that we performed resulted in significant associations is not surprising: many, if not most, markers genome-wide are expected to be correlated with harvest date and firmness since the genetic structure of our apple population is strongly correlated with these traits. For example, without correcting for the effects of population structure, 39% and 17% of genome-wide SNPs were significantly associated ($P < 0.05$) with harvest date and firmness, respectively, in the USDA collection, which is genetically identical to most of the population studied here (Migicovsky et al, 2016). Because we performed single-marker association tests in the present study, we were unable to account for population
structure using genome-wide markers as is customarily done when performing GWAS. Thus, we
were only able to compare the relative power of each marker to predict firmness and harvest
date. We found that the D5Y marker in \textit{NAC18.1} had a 3 to 14 times greater effect on firmness
and harvest than the markers in \textit{ACO1}, \textit{ACS1} and \textit{PG1} (Figure 1). The superior predictive power
of the NAC18.1 marker suggests that it has more potential to advance genomics-assisted
breeding for firmness and harvest date compared to markers currently in use. Results from our
forthcoming GWAS that include all of these markers and properly account for the effects of
relatedness and population structure will be necessary to definitively conclude which, if any, of
these traditionally used markers predict firmness and/or harvest date in diverse collections of
apple.

While the D5Y mutation in \textit{NAC18.1} is a strong functional candidate variant, we
discovered numerous DNA sequence variants in NAC18.1, often in perfect LD with D5Y, that
are also putatively functional (Figure 2). The magnitude of DNA polymorphism at this locus,
and the multitude of putatively causal variants, are consistent with the results of Larsen et al.
(2019), who found 18 SNPs and 2 indels within \textit{NAC18.1} across 11 apple accessions. Given the
pattern of LD we observed across \textit{NAC18.1} (Figure 2), we considered the possibility that the
GWAS signals at \textit{NAC18.1} are in fact driven by a causal variant outside of the \textit{NAC18.1} coding
sequence that acts independently of the \textit{NAC18.1} gene. Since \textit{NAC18.1} is a homolog of the well-
known ripening gene \textit{NOR} in tomato, we used tomato as a model to explore the role of \textit{NAC18.1}
in the ripening process.

We observed enhanced ripening in the \textit{nor} mutant following heterologous expression of
either of the \textit{NAC18.1} alleles. Notably, the synthesis of carotenoids was restored in the
transgenic fruit, resulting in visual color change that approached the level of WT at the fruit
surface (Figure 3B), although dissection of fruit revealed incomplete internal color change in all lines (Figure 3A). Visual color change was consistent with the expression of the carotenoid biosynthetic gene \textit{PSY1} (Figure 4B). Gene expression analysis further confirmed the induction of genes involved in ripening-associated cell wall remodeling (Figure 4C) and ethylene synthesis (Figure 4D). Although the expression level of the \textit{NAC18.1} transgene was comparable across transgenic lines (Figure 4A), the level of ripening marker induction in the distinct lines was significantly different (Figure 4B-D). We had hypothesized that an allele-specific trend in our transgenic experiments would indicate that coding sequence variants are responsible for the association of \textit{NAC18.1} with variation in apple ripening. While our results do not indicate such a trend, they do confirm that both alleles of \textit{NAC18.1} encode functional proteins that can promote ripening.
Figure 4. Ripening marker gene expression in transgenic tomatoes. Gene expression was evaluated as the difference in threshold cycle (ΔCt, log2 scale) by qRT-PCR using RPL2 as a reference gene. The mean ± SE is superimposed in black over the raw values in gray (n = 5). Genotypes not sharing a letter (a-d) are statistically distinct by one-way ANOVA and Tukey’s HSD test (p < 0.05). A. The NAC18.1 transgene. B. PSY1, encoding the phytoene synthase carotenoid biosynthetic gene. C. PG2, encoding the major ripening-associated polygalacturonase involved in pectin hydrolysis. D. ACS2, encoding 1-aminocyclopropane-1-carboxylate synthase, a ripening-associated isoform of the ethylene biosynthesis pathway.

In tomato, nor was first identified as a spontaneous mutation in an heirloom cultivar (Tigchelaar et al., 1973). While this nor allele exhibits recessive behavior and has been assumed to be a complete loss-of-function, due to a 2 nt deletion resulting in a truncated protein (Giovannoni et al., 2004), it was recently demonstrated that it is actually a dominant negative allele (Wang et al., 2019; Gao et al., 2019). These studies used CRISPR/Cas9 to generate bona fide null alleles of nor, which showed evidence of ripening relative to the spontaneous nor allele, although to a lesser extent than WT fruit. In light of these studies, we speculate that the action of the NAC18.1 transgenes was likely attenuated by the dominant-negative activity of the spontaneous nor allele used in our work. Further heterologous characterization of NAC18.1 would likely benefit from the use of null nor mutants, or double mutants of nor and nor-like 1 (Gao et al., 2018). This may allow for more precise quantitative comparisons between alleles in order to resolve whether differences in the coding sequence of NAC18.1 confer different degrees of ripening.
Although we are unable to conclude whether polymorphisms in coding or regulatory sequences affect the activity of NAC18.1 in apple, an increasing body of evidence generated using tomato as a model system indicates that coding sequence polymorphisms of NOR can influence firmness and timing of ripening. For example, the Portuguese Alcobaça cultivar of tomato has firm fruit, delayed ripening, and long shelf life that is conferred by the alcobaça mutation, and a recent study has shown that ALCOBACA is allelic with NOR, and that the alcobaça allele of NOR contains a valine to aspartate mutation at position 106 within the NAC domain (Kumar et al., 2018). An additional complete loss of function allele of NOR was also found in another long shelf life tomato cultivar (Kumar et al., 2018).

Our analysis of NAC18.1 sequences revealed a number of mutations that are candidates for causal association with ripening phenotypes. For example, we observed an insertion of four amino acids (QPQP) 11 amino acids downstream of the D5Y mutation in all A haplotypes upstream of the NAC DNA binding domain. Glutamine-rich sequences are common in eukaryotic transcription factors, and polymorphisms in these motifs have been shown to alter the activity of transcriptional activators (Atanesyan et al., 2012). On the other hand, we also observed several SNPs resulting in amino acid changes within the DNA binding domain and C-terminal transcriptional activator domain of NAC18.1. Clearly, resolution of a causal polymorphism within the coding sequence or promoter of NAC18.1 will require substantial additional effort. In this regard, we are resequencing the NAC18.1 locus in our complete diversity panel in order to identify intragenic recombination that is informative in combination with our long-term phenotyping of the collection.

In conclusion, the results presented here provide evidence of the superiority of GWAS for the identification of genetic markers associated with firmness and harvest date in diverse
germplasm. By validating NAC18.1 as a functional candidate for apple fruit ripening, we lay the groundwork for future efforts to compare the effect of different NAC18.1 haplotypes and determine the precise causal variant(s) underlying this agriculturally important gene. Ultimately, we envision that the identification of precise causal variants for ripening in apple could be exploited using gene editing for fast and efficient introgression of this desirable trait.

Acknowledgments

We thank Susan Brown (Cornell University), Benjamín Gutiérrez (USDA-ARS Plant Genetics Resource Unit), and C. Thomas Chao (USDA-ARS Plant Genetics Resource Unit) for apple leaf tissue samples from orchards of the Cornell apple breeding program and the USDA apple germplasm collection. Transgenic tomato plants were generated by the Boyce Thompson Institute Biotechnology Center (https://btiscience.org/). ZM was supported by National Science Foundation Plant Genome Research Program 1546869. TY and JR were supported by a grant from the USDA-ARS Multistate (NE-1336) project, NYC-184821. This research was supported in part by funding the National Sciences and Engineering Research Council of Canada (SM) and A-Base funding (NOI-1238) from Agriculture and Agri-Food Canada (JS). The authors gratefully acknowledge the farm services team at the Kentville Research and Development Centre for maintaining the apple collection studied here.
Supplemental Figure Legends

Supplemental Figure S1. Complete amino acid sequence alignment of 24 NAC18.1 haplotypes.

Supplemental Table Legends

Supplemental Table S1. PCR primers used in this work.

Supplemental Table S2. Genotypes of NAC18.1 (D5Y), ACS1, and ACO1 across 866 apple accessions.
References

Atanesyan L, Günther V, Dichtl B, Georgiev O, Schaffner W. 2012. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. Biological Chemistry 393, 63–70.

Atkinson RG, Sutherland PW, Johnston SL, Gunaseelan K, Hallett IC, Mitra D, Brummell DA, Schröder R, Johnston JW, Schaffer RJ. 2012. Down-regulation of POLYGALACTURONASE1 alters firmness, tensile strength and water loss in apple (Malus x domestica) fruit. BMC Plant Biology 12, 129.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67, 1–48.

Biggs MS, Handa AK. 1989. Temporal Regulation of Polygalacturonase Gene Expression in Fruits of Normal, Mutant, and Heterozygous Tomato Genotypes. Plant Physiology 89, 117–125.

Blanpied GD, Silsby KJ. 1992. Predicting Harvest Date Windows for Apples.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter 11, 113–116.

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, Van de Weg WE. 2010. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus × domestica Borkh.). Journal of Experimental Botany 61, 3029–3039.
Costa F, Stella S, Van de Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S. 2005. Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus domestica Borkh). Euphytica 141, 181–190.

Curtis MD, Grossniklaus U. 2003. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. Plant Physiology 133, 462–469.

Daccord N, Celton J-M, Linsmith G, et al. 2017. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics 49, 1099–1106.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Fray RG, Grierson D. 1993. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Molecular Biology 22, 589–602.

Gao Y, Wei W, Zhao X, et al. 2018. A NAC transcription factor, NOR-like1, is a new positive regulator of tomato fruit ripening. Horticulture Research 5, 75.

Gao Y, Zhu N, Zhu X, et al. 2019. Diversity and redundancy of the ripening regulatory networks revealed by the fruitENCODE and the new CRISPR/Cas9 CNR and NOR mutants. Horticulture Research 6, 39.

García-Gómez BE, Salazar JA, Dondini L, Martínez-Gómez P, Ruiz D. 2019. Identification of QTLs linked to fruit quality traits in apricot (Prunus armeniaca L.) and biological validation through gene expression analysis using qPCR. Molecular Breeding 39, 28.
Giovannoni J, Nguyen C, Ampofo B, Zhong S, Fei Z. 2017. The Epigenome and Transcriptional Dynamics of Fruit Ripening. Annual Review of Plant Biology 68, 61–84.

Giovannoni J, Tanksley S, Vrebalov J, Noensie F. 2004. NOR gene compositions and methods for use thereof.

Harada T, Sunako T, Wakasa Y, Soejima J, Satoh T, Niizeki M. 2000. An allele of the 1-aminocyclopropane-1-carboxylate synthase gene (Md-ACS1) accounts for the low level of ethylene production in climacteric fruits of some apple cultivars. Theoretical and Applied Genetics 101, 742–746.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33, 1870–1874.

Kumar R, Tamboli V, Sharma R, Sreelakshmi Y. 2018. NAC-NOR mutations in tomato Penjar accessions attenuate multiple metabolic processes and prolong the fruit shelf life. Food Chemistry 259, 234–244.

Larsen B, Migicovsky Z, Jappesen AA, Gardner KM, Toldam-Andersen TB, Myles S, Ørgaard M, Petersen MA, Pedersen C. 2019. Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition and Harvest Date. The Plant Genome in press.

Lenth RV. 2016. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software 69, 1–33.
Lü P, Yu S, Zhu N, et al. 2018. Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. Nature Plants 4, 784.

Luby JJ, Shaw DV. 2001. Does Marker-assisted Selection Make Dollars and Sense in a Fruit Breeding Program? HortScience 36, 872–879.

Mathew IE, Agarwal P. 2018. May the Fittest Protein Evolve: Favoring the Plant-Specific Origin and Expansion of NAC Transcription Factors. BioEssays 40, 1800018.

McClure KA, Gardner KM, Douglas GM, et al. 2018. A Genome-Wide Association Study of Apple Quality and Scab Resistance. The Plant Genome 11, 0.

Migicovsky Z, Gardner KM, Money D, et al. 2016. Genome to Phenome Mapping in Apple Using Historical Data. The Plant Genome 9, 0.

Migicovsky Z, Li M, Chitwood DH, Myles S. 2018. Morphometrics Reveals Complex and Heritable Apple Leaf Shapes. Frontiers in Plant Science 8.

Nakatsuka A, Murachi S, Okunishi H, Shiomi S, Nakano R, Kubo Y, Inaba A. 1998. Differential Expression and Internal Feedback Regulation of 1-Aminocyclopropane-1-Carboxylate Synthase, 1-Aminocyclopropane-1-Carboxylate Oxidase, and Ethylene Receptor Genes in Tomato Fruit during Development and Ripening. Plant Physiology 118, 1295–1305.

Oraguzie NC, Iwanami H, Soejima J, Harada T, Hall A. 2004. Inheritance of the Md-ACS1 gene and its relationship to fruit softening in apple (Malus × domestica Borkh.). Theoretical and Applied Genetics 108, 1526–1533.
Ornelas-Paz J de J, Quintana-Gallegos BM, Escalante-Minakata P, Reyes-Hernández J, Pérez-Martínez JD, Rios-Velasco C, Ruiz-Cruz S. 2018. Relationship between the firmness of Golden Delicious apples and the physicochemical characteristics of the fruits and their pectin during development and ripening. Journal of Food Science and Technology 55, 33–41.

Osorio S, Alba R, Damasceno CMB, et al. 2011. Systems Biology of Tomato Fruit Development: Combined Transcript, Protein, and Metabolite Analysis of Tomato Transcription Factor (nor, rin) and Ethylene Receptor (Nr) Mutants Reveals Novel Regulatory Interactions. Plant Physiology 157, 405–425.

Pirona R, Eduardo I, Pacheco I, et al. 2013. Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. BMC Plant Biology 13, 166.

Shinozaki Y, Nicolas P, Fernandez-Pozo N, et al. 2018. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nature Communications 9, 364.

Solé X, Guinó E, Valls J, Iniesta R, Moreno V. 2006. SNPStats: a web tool for the analysis of association studies. Bioinformatics 22, 1928–1929.

Sunako T, Sakuraba W, Senda M, Akada S, Ishikawa R, Niizeki M, Harada T. 1999. An allele of the Ripening-Specific 1-Aminocyclopropane-1-Carboxylic Acid Synthase Gene (ACS1) in Apple Fruit with a Long Storage Life. Plant Physiology 119, 1297–1304.

Tigchelaar EC, Tomes ML, Kerr EA, Barman RJ. 1973. A new fruit ripening mutant, non-ripening (nor). Rep. Tomato Genet. Coop 23, 33.
Urrestarazu J, Muranty H, Denancé C, et al. 2017. Genome-Wide Association Mapping of Flowering and Ripening Periods in Apple. Frontiers in Plant Science 8.

Van Eck J, Keen P, Tjahjadi M. 2019. Agrobacterium tumefaciens-Mediated Transformation of Tomato. In: Kumar S, Barone P, Smith M, eds. Transgenic Plants: Methods and Protocols. New York, NY: Springer New York, 225–234.

Wang R, Tavano EC da R, Lammers M, Martinelli AP, Angenent GC, Maagd RA de. 2019. Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Scientific Reports 9, 1696.

Wang D, Yeats TH, Uluisik S, Rose JKC, Seymour GB. 2018. Fruit Softening: Revisiting the Role of Pectin. Trends in Plant Science 23, 302–310.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.