Cell type-specific control of secondary cell wall formation by Musashi-type translational regulators in Arabidopsis

Alicia Kairouani,[1]* Dominique Pontier,[1]* Claire Picart,[1] Fabien Mounet,[2] Yves Martinez,[3] Lucie Le-Bot,[4] Mathieu Fanuel,[4,5] Philippe Hammann,[6] Lucid Belmudes,[7] Rémy Merret,[1] Jacinthe Azevedo,[1] Marie-Christine Carpentier,[1] Dominique Gagliardi,[8] Yohann Couté,[7] Richard Sibout,[4] Natacha Bies-Etheve,[1@] Thierry Lagrange[1@]

[1]Laboratoire Génome et Développement des Plantes, Université de Perpignan via Domitia, CNRS, UMR5096, Perpignan Cedex, France.

[2]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse III, CNRS UPS UMR5546, Castanet-Tolosan, France.

[3]FRAIB-CNRS plateforme imagerie, Castanet-Tolosan, France.

[4]Biopolymères Interactions Assemblages, UR1268 BIA, INRAE, Nantes, France.

[5]PROBE research infrastructure, BIBS Facility, INRAE, Nantes, France.

[6]Plateforme Protéomique Strasbourg Esplanade de CNRS, Université de Strasbourg, Strasbourg, France.

[7]Université Grenoble Alpes, INSERM, UA13 BGE, CNRS, CEA, FR2048, Grenoble, France.

[8]Institut de Biologie Moléculaire des Plantes, BMP, CNRS, Université de Strasbourg, Strasbourg, France.

*These authors contributed equally to this work.

@ Correspondence to: Natacha Bies-Etheve, Thierry Lagrange (ORCiD: 0000-0003-3090-0918).

Email : natacha.etheve@univ-perp.fr; lagrange@univ-perp.fr

Deciphering the mechanism of secondary cell wall/SCW formation in plants is key to understanding their development and the molecular basis of biomass recalcitrance. Although transcriptional regulation is essential for SCW formation, little is known about the implication of post-transcriptional mechanisms in this process. Here we report that two *bonafide* RNA-binding proteins homologous to the animal translational regulator Musashi, MSIL2 and MSIL4, function redundantly to control SCW formation in Arabidopsis. MSIL2/4 interactomes are similar and enriched in proteins involved in mRNA binding and translational regulation. MSIL2/4 mutations alter SCW formation in the fibers, leading to a reduction in lignin deposition, and an increase of 4-*O*-glucuronoxylan methylation. In accordance, quantitative proteomics of stems reveal an overaccumulation of glucuronoxylan biosynthetic machinery, including GXM3, in the *msil2/4* mutant stem. We showed that MSIL4 immunoprecipitates *GXM* mRNAs, suggesting a novel aspect of SCW regulation, linking translational control to the regulation of SCW biosynthesis genes.

50 **Introduction**

51 As the main constituent of the terrestrial plant biomass, lignocellulose represents a huge
52 reservoir of fixed carbon and a renewable resource for bioproducts and energy that is essential
53 for human usages[1]. Lignocellulose biomass enclosed mainly within secondary cell wall
54 (SCW) reinforces vessels for long-distance transport and provides physical properties to
55 fibers allowing upright growth of the plants[1,2]. SCWs are composite material made of high
56 molecular weight biopolymers, including cellulose, hemicelluloses, and lignin, as well as cell
57 wall proteins whose composition and structure can differ markedly between dicots and
58 monocots[2,3]. Cellulose is synthesized at the plasma membrane by the cellulose synthase
59 complex and is composed of linear chains of β-(1–4) glucans that aggregate to form highly
60 crystalline cellulose microfibrils[2,3]. Xylans, the main hemicellulosic polysaccharides in SCW,
61 are synthesized in the Golgi apparatus by a complex biosynthetic machinery and comprise a
62 group of polysaccharides that share a common backbone of β-(1,4)-linked xylose (Xyl) units,
63 but differs by the presence of side chains, whose nature is dependent on tissues and species[2].
64 In particular, glucuronoxylan, the dominant xylan in the SCW of dicots, is notably substituted
65 at *O*-2 position with α-D-glucuronic acid (GlcA) or (4-O-methyl)- α-D-glucuronic acid
66 (MeGlcA) groups, and with O-acetyl moieties at C-2 or C-3 positions[2]. By interacting with
67 the cellulose and lignin, xylans contribute to the strengthening of the SCW, and are
68 considered to be one of the main factors contributing to the biomass recalcitrance to
69 enzymatic hydrolysis[3,4]. Finally, lignin polymers are made up of H, G, and S units which are
70 the results of the oxidation and the polymerisation of the *p*-coumaryl, coniferyl and synapyl
71 alcohols, respectively. These monomers, also called monolignols are synthesized on the ER
72 surface/cytoplasm prior transportation to the cell wall. Lignins confer stiffness, strength and
73 hydrophobicity to the SCW[2-4].

74 The metabolic load imposed by the coordinated production of SCW biopolymers
75 emphasizes the importance for plants of having a precise control over temporal and spatial
76 expression of the corresponding SCW biosynthetic genes. Based on founding studies
77 performed in Arabidopsis model, it is generally recognized that the SCW synthesis is
78 primarily controlled at a transcriptional level by a multilayer and interconnected network of
79 transcription factors/TFs[5,6]. Key players of this regulatory network are related NAC (NAM-
80 ATAF1/2-CUC) and MYB-type TFs that induce, in a redundant and combinatorial manner,
81 the expression of SCW biosynthetic genes by binding their promoter region[5]. In addition, a
82 family of Class III homeodomain leucine zipper (HD-ZIPIII) TFs, whose members control
83 several aspects of Arabidopsis development, was also shown to interact with SCW regulatory
84 network, contributing to xylem cell specification and SCW synthesis[5,6]. Although this
85 transcriptional level of control can gain complexity in perennial plants, functional orthologs
86 of the main regulators were identified in many different species, suggesting that SCW
87 deposition proceeds via a conserved mechanism in all vascular plants[7,8]. During the last
88 decade, engineering the SCW regulatory network to deposit modified lignin and/or improve
89 polysaccharide composition became a promising strategy to optimize biomass
90 processability[9,10]. However, despite huge progresses in understanding the regulation of SCW
91 synthesis, the engineering of SCW regulatory network to improve biomass yield and
92 digestibility has proven difficult and often impairs plant growth because of altered vascular
93 tissue development[11,12].

94 Downstream of transcriptional regulations by TFs, RNA-binding proteins (RBPs) are
95 essential post-transcriptional modulators of gene expression across all kingdoms of life[13].
96 However, although a large cohort of RBPs exists in plants[14], there is little information about
97 RBPs-mediated posttranscriptional regulation of SCW biosynthetic genes. So far, only few
98 microRNAs (miRNAs) families have been implicated in the post-transcriptional regulation of

99 genes involved in both regulatory and enzymatic aspects of SCW biosynthesis[6]. In particular,
100 miR165/166 were shown to target HD-ZIPIII transcriptional regulators, having a direct
101 impact on the transcriptional network associated with SCW biosynthesis[6]. In the other hand,
102 miRNA397/857/408 were shown to affect lignin content *via* the targeting of laccase/LAC
103 genes, which control lignin polymerization from monolignols precursors[15,16]. In addition, two
104 related tandem CCCH zinc-finger proteins, which exhibit both DNA and RNA binding
105 abilities, have been proposed to modulate SCW formation by regulating the expression of
106 genes associated with cell wall metabolism[17]. Despite these mounting evidences, our current
107 knowledge of the role of post-transcriptional regulators in SCW synthesis is still very limited.

108 In this work, we report that two RRM-domain containing RNA-binding proteins
109 homologous to the animal translational regulator Musashi, Musashi-like2/MSIL2 and
110 Musashi-like4/MSIL4, function redundantly to control various aspects development,
111 including the stiffness of the inflorescence stem in Arabidopsis. We show that RRM-
112 dependent RNA-binding activity is essential for MSIL2/4 functions *in vivo*, and that MSIL2/4
113 interactomes are similar, being enriched in proteins involved in 3'-UTR binding and
114 translational regulation. MSIL2/4 mutations alter the formation of SCW in the interfascicular
115 fiber cells, leading to a reduction in lignin deposition, and a change in the decoration pattern
116 of glucuronoxylan that is associated with an increase of 4-*O*-methylation of GlcA substituent.
117 In accordance, quantitative mass-spectrometry-based protein analysis reveal an
118 overaccumulation of glucuronoxylan biosynthetic machinery, including the GlucuronoXylan
119 Methyltransferase3/GXM3, in the *msil2/4* mutant stem. We show that MSIL4
120 immunoprecipitates GXM3 mRNA in vivo, likely regulating its expression at a translational
121 level. Our results demonstrate that MSILs regulate SCW synthesis in interfascicular fiber
122 cells and point to a novel aspect of SCW regulation linking translational repression to
123 regulation of SCW biosynthesis genes.

124 **Results**

125 **Musashi-like MSIL2/4 proteins redundantly control specific Arabidopsis development**
126 **processes.**

127 While many studies have implicated the MSIs in the control of gene expression during
128 cellular proliferation, cell fate determination and cancer in animals[18,19], yet little is known
129 about the presence and activity of MSI-type proteins in plants. In this work, we describe an
130 RNA-binding protein family in Arabidopsis, whose members hereafter named as *MUSASHI-*
131 *Like1* to *4/MSIL1-4*, share both sequence, domain organization, and model-based structural
132 similarities with animal MSIs (Fig. 1a, b). Phylogenetic analysis revealed that Arabidopsis
133 indeed harbors a clade of seven *MSIL*-type genes with a prominent sub-clade harboring the
134 *MSIL1-4* genes (Fig. 1b). Notably, a neighboring sub-clade contains two genes, *RBGD2* and
135 *RGBD4* (Fig. 1b), that encode for heat-inducible RBPs implicated in the response to heat
136 stress in Arabidopsis[20]. MSIL1-4 proteins share a common domain architecture consisting of
137 two N-terminal RNA recognition motifs (RRM1 and RRM2) that are followed by a poorly
138 conserved, intrinsically unstructured, carboxy-terminal region (Fig. 1a; and Supplementary
139 Fig. 1a). The survey of the gene expression ARAPORT11 databases (https://araport.org/;
140 http://fgcz-pep2pro.uzh.ch) revealed that the *MSIL1-4* genes are widely expressed in
141 Arabidopsis, with *MSIL1* showing the lowest levels of expression (Fig. 1c). Database searches
142 further indicated that *MSIL* orthologs are widely distributed in land plants, from bryophytes to
143 angiosperms (Supplementary Fig. 1b). To gain a mechanistic understanding of MSIL function
144 in Arabidopsis, we determined the subcellular localization of these proteins by performing
145 fluorescence microscopy on GFP-tagged versions of MSIL2 and MSIL4 (MSIL2G and
146 MSIL4G). Both proteins showed a diffuse cytoplasmic distribution in root cells of stable
147 Arabidopsis transgenic plants that contrasts with the nucleus and cytoplasmic distributions of
148 the free GFP protein control (Fig. 1d).

149 To further investigate the function of the MSILs in Arabidopsis, we raised specific
150 antibodies against non-conserved regions of MSIL1-4 and characterized T-DNA insertion
151 *msil* mutant lines that lack full-length *MSIL* mRNA and MSIL proteins as judged by RT-PCR
152 and western blot experiments, respectively (Supplementary Fig. 2a-c). However, none of the
153 four characterized single *msil* mutants showed any obvious phenotype (Supplementary Fig. 3a
154 and data not shown). To assess the potential functional redundancy between *MSIL* genes, we
155 crossed the single *msil* mutants together and characterized the six possible double mutant
156 combinations (Supplementary Fig. 3a). Only the *msil2*-1/*msil4*-1 (*msil2/4*) double mutant
157 showed discernable developmental abnormal phenotypes, including enlarged and curled
158 leaves, early leaf senescence and a pendant inflorescence stem (Fig. 1e, f; and Supplementary
159 Fig. 3a, b). To assess whether MSIL1 and MSIL3 also contribute to Arabidopsis development
160 in the *msil2/4* double mutant background, we generated a *msil1*-1/*msil2*-1/*msil3*-1/*msil4*-1
161 (*msil1/2/3/4*) quadruple mutant that shows no particular or aggravated developmental
162 phenotypes with respect to the *msil2/4* double mutant (Supplementary Fig. 3a, c, d). To test
163 the functional redundancy of MSIL2 and MSIL4 proteins *in vivo*, we transformed the *msil2/4*
164 double mutant with constructs coding either for a Flag/HA-tagged version of MSIL2
165 (MSIL2F) or MSIL4 (MSIL4F), and selected two independent transgenic plants expressing
166 similar and near physiological levels of the tagged protein, as evaluated by western blot using
167 our home-made antibodies (Supplementary Fig. 3e). The ectopic expression of either MSIL2F
168 or MSIL4F proteins was able to rescue all the phenotypes associated with the MSIL2/4 defect
169 (Fig. 1e, f; and Supplementary Fig. 3b). This confirms that MSIL2F and MSIL4F are
170 functional proteins and more importantly that MSIL2 and MSIL4 redundantly control various
171 aspects of development in Arabidopsis.
172
173 **RRM-dependent RNA-binding activity is essential for MSIL2/4 functions *in vivo*.**

174 Animal MSIs interact with target transcripts through their RRM domains and mutations in
175 these domains have been shown to impair MSI activity *in vivo*[21,22]. MSILs are very similar to
176 metazoan MSIs over their RRM sequences (Fig. 1a; and Supplementary Fig. 4a), a closeness
177 of sequences also observed when homology models of the MSIL4 RRM motifs were
178 generated using SWISS-MODEL server[23] (Fig. 2a). Homology-based structure prediction
179 proposed that these domains adopt the characteristic RRM fold with a four-stranded β-sheet
180 structure bearing the conserved phenylalanine residues involved in the specific recognition of
181 RNA bases[24] (Fig. 2a; and Supplementary Fig. 4a). This supports the idea that MSILs have
182 retained the ability to interact with RNA, a notion further supported by the identification of
183 MSIL2/4 proteins in the experimentally determined Arabidopsis mRNA-binding proteome[14]
184 To examine the requirement for the interaction of MSIL with RNA in developmental control,
185 we first generated an RNA-binding mutant form of MSIL4, MSIL4^{RRM}, by introducing
186 phenylalanine to aspartate (F→D) mutations in both RRM domains (Fig. 2b). Protein-RNA
187 interaction assays using RNA homopolymers immobilized on agarose beads confirmed that
188 the MSIL4, but not MSIL4^{RRM}, exhibited intrinsic RNA binding activity *in vitro* (Fig. 2c). To
189 assess the impact of these mutations *in vivo*, the MSIL4G and MSIL4G^{RRM} constructs were
190 expressed under the control of the *MSIL4* endogenous promoter into the *msil2/4* mutant, and
191 two independent transformants expressing near physiological levels of the WT (*MSIL4G*-3
192 and *MSIL4G*-8), and mutant (*MSIL4G^{RRM}*-3 and *MSIL4G^{RRM}*-10) proteins were selected for
193 further complementation analysis (Supplementary Fig. 4b). MSIL4G, but not MSIL4G^{RRM}
194 proteins were able to rescue the developmental defects incurred by *MSIL2/4* mutations (Fig.
195 2d; and Supplementary Fig. 5a, b), indicating that these proteins are likely to exert their
196 functions by interacting with RNA targets *in vivo*.
197

198 **MSIL2/4 protein interactomes are enriched in proteins involved in 3'-UTR binding and**
199 **translational regulation.**

200 Animal MSIs modulate mRNA expression mostly at a translational level through binding to
201 conserved motifs in the 3'-UTR of the target mRNA and further interactions with various
202 mRNA binding partners[18,25,26]. To further characterize the components of the MSIL2/4
203 network in Arabidopsis, we performed affinity purification coupled to LC-MS/MS, as
204 described in Scheer et al.[27], using the complemented *MSIL2F1* and *MSIL4F1* lines. Consistent
205 with their functional redundancy, MSIL2F1 and MSIL4F1 exhibit a similar protein interaction
206 network, with 27 proteins significantly enriched in both MS analyses (Fig. 3a, b; and
207 Supplementary Table 1). Examination of the PANTHER database (http://www.pantherdb.org)
208 identified that MSIL2F1/4F1 are part of a protein-protein interaction network that was
209 principally enriched in 4 categories of GO molecular functions, namely poly(A) binding
210 protein (GO:0008143), mRNA 3-UTR binding (GO:0003730), single-stranded RNA binding
211 (GO:0003727), and translation initiation factor activity (GO:0003743), highlighting the
212 relationship between MSIL2/4 and the mRNA binding and regulatory proteome (Fig. 3c).

213 Many of the proteins present in the MSIL2/4 interactome, such as PABs, PUMs,
214 UBP1 are involved in translation regulation and are targeted to stress granule (SGs)
215 components, a membraneless cytoplasmic organelle that formed consecutive to the global
216 inhibition of translation caused by heat-stress[28,29]. To assess the potential functional link
217 between MSIL and those proteins, we crossed the MSIL2G/4G plants with a functional
218 PAB2-RFP/PAB2R line[29], and assessed the localization of the fusion proteins before and after
219 heat-stress in both parent and crossed lines. Fluorescence analysis revealed that MSIL2G/4G,
220 like PAB2R, show a dispersed cytoplasmic signal under normal condition, and are recruited
221 to cytoplasmic granules that overlap with PAB2R-containing SG upon heat stress (Fig. 3d;
222 and Supplementary Fig. 6). Treatment with cycloheximide (CHX), a translation inhibitor that

223 prevent SG assembly by trapping mRNAs at polysomes[30], confirmed that those foci
224 correspond to *bonafide* SGs (Fig. 3e). These results were consistent with the RNA-binding
225 protein-based signature and the nature of MSIL interactomes, further supporting a functional
226 relationship between MSILs and the translation machinery in Arabidopsis.
227
228 **MSIL2/4 proteins regulate the molecular architecture of SCW in Arabidopsis fibers.**
229 Pendant stem inflorescence is a common phenotypic trait of Arabidopsis mutants defective in
230 SCW biosynthesis, that results from a lack of mechanical support and rigidity[6,31]. Given the
231 current lack of knowledge of the role of post-transcriptional mechanisms in the control of
232 SCW biosynthesis in plants, we henceforth focused our analysis on the pendant stem
233 phenotype of the *msil2/4* double mutant. Survey of the Arabidopsis inflorescence stem tissue-
234 specific transcriptome database[32] indicated that the *MSIL2/4* genes are highly expressed in the
235 SCW-forming xylary fiber (F) and xylem vessel (X) cells (Fig. 4a). Toluidine blue O staining
236 and microscopic analysis of Col-0 and *msil2/4* mutant stem cross-sections revealed no major
237 changes in the architecture of both the interfascicular and xylary fibers in the mutant stem.
238 Notably, xylem cells deformation or collapse, which are a characteristic of mutants deficient
239 in SCW synthesis, were not observed in the *msil2/4* mutant (Fig. 4b). However, the
240 examination of the cell wall thickness in cross sections of stems revealed that the
241 interfascicular fiber cells display a significantly thinner SCW in *msil2/4* mutant (Fig. 4c). UV
242 autofluorescence and phloroglucinol staining of the Col-0, *msil2/4*, *msil2/4+MSIL2*F1 and
243 *msil2/4+MSIL4*F1 stem sections revealed a reduced lignin signal in the interfascicular fibers of
244 the *msil2/4* mutant compared to Col-0 or complemented plants (Fig. 4d; and supplementary
245 Fig. 7a). In agreement with these observations, a significant decrease of 30% in AcBr lignin
246 content was observed in the *msil2/4* mutant compared to the Col-0 control plants (Fig. 4e).
247 Despite a tendency to decrease, the difference of lignin content was not significant in the

248 complemented lines *msil2/4+MSIL2*F1 and *msil2/4+MSIL4*F1 compared to Col-0 (Fig. 4e). To
249 further investigate potential changes in the lignin structure and composition in the Col-0 and
250 *msil2/4* mutant, we evaluated the impact of MSIL2/4 deficiency on lignin composition by
251 thioacidolysis, a degradation method allowing the determination of the relative amounts of H,
252 G and S lignin units linked by beta-O-4 linkages. Total monomer yield did not significantly
253 differ between the col-0 and the *msil2/4* mutant (Supplementary Fig. 7b), suggesting that the
254 structure of the lignin polymer is not merely affected in the mutant. Further examination of β-
255 glucan composition using the fluorescent dye calcofluor white[33] revealed an increased staining
256 in the interfascicular fiber cells of *msil2/4* mutant, supporting changes in the
257 composition/accessibility of cellulose (Fig. 4f; and Supplementary Fig. 7c).

258 To extend this analysis, we then subjected purified Col-0 and *msil2/4* SCW materials
259 to Fourier-Transform Infrared (FTIR) spectroscopy[34]. PLS-DA analyses performed on FT-IR
260 absorbance data clearly separated *msil2/4* mutant lines from Col-0 plants, using the first two
261 components (explaining 74% and 10% of total variability, respectively) (Supplementary Fig.
262 7d, left panel). The large-scale chemotyping of CW composition in stem by FTIR
263 spectroscopy suggested that not only lignin deposition was impaired in *msil2/4* mutants but
264 also polysaccharides composition (Supplementary Fig. 7d, right panel). The lowest
265 absorbance values in *msil2/4* lines were observed in the region 1800-1500cm^{-1}, mostly
266 associated to the absorbance of lignin related compounds. On the contrary, a higher
267 absorbance was detected in *msil2/4* samples within 1050-900 cm^{-1} region, mostly associated
268 to polysaccharides compounds[34]. Because the optimal interaction between the SCW lignin and
269 polysaccharide polymers is believed to contribute to SCW recalcitrance, we used
270 saccharification as a proxy to compare cell wall sugars accessibility between Col-0, *msil2/4*
271 and complemented mutant lines (Fig. 4g). A higher amount of reducing sugars from native
272 cell wall (without pre-treatment/-Tr) was released using *msil2/4* samples compared to Col-0

273 samples. After cell wall loosening using alkali pre-treatment/+Tr, an even more important
274 amounts of sugars (+5%) was released from double mutants CW enzymatic digestion (Fig.
275 4g). Together, our data indicate that MSIL2/4 regulate the molecular architecture of SCW in
276 the interfascicular fibers, contributing significantly to the setting of biomass recalcitrance in
277 Arabidopsis.
278
279 **The accumulation and activity of the glucuronoxylan decoration machinery are mis-**
280 **regulated in *msil2/4* mutant.**
281 To elucidate the impact of the MSIL2/4 on the expression of SCW-related biosynthesis genes
282 in the Arabidopsis inflorescence stem, we performed mRNA sequencing (mRNA-seq) on both
283 wild-type and *msil2/4* mutant backgrounds. In total, we found 234 genes to be significantly
284 differentially regulated (DEG) in the *msil2/4* mutant background (with stringent thresholds
285 having a logarithm of fold change >1.5 or <-1.5 and false discovery rate (FDR ≤ 0.05),
286 including 156 up-regulated and 78 down-regulated mRNAs (Supplementary Fig. 8a; and
287 Supplementary Table 2). Interestingly, genes encoding regulatory or enzymatic components
288 of the SCW biosynthesis pathway were absent from the DEG list, and their expression was
289 not significant changed in *msil2/4* mutant (Supplementary Fig. 8a). Gene Ontology analysis of
290 DEGs revealed significantly enriched signaling pathways (FDR ≤0.05) involved in plant
291 defense or responses to biotic and abiotic stresses (Supplementary Fig. 8b), encoding products
292 that are preferentially targeted to the extracellular region and the cell wall (Supplementary
293 Fig. 8c). In this regard, the constitutive activation of defense pathways has been well
294 documented in mutant plants defective in cell wall formation, including the SCW[35].
295 The absence of SCW biosynthesis genes in the DEG list suggests that MSIL2/4 could
296 regulate the SCW formation at a post-transcriptional/translational level, a hallmark of MSI
297 function in animal models[18,19]. To investigate proteome-level changes that could potentially

298 occur upon MSIL2/4 mutations, we performed a comparative proteomic analysis of Col-0 and
299 *msil2/4* inflorescence stems using MS-based quantitative proteomics, and represented ≈4300
300 reliably identified and quantified proteins in a volcano plot according to their statistical p-
301 value and their relative difference of abundance. This analysis revealed 267 proteins with
302 significant abundances in the *msil2/4* mutant versus Col-0 inflorescence stems (fold change
303 ≥2, Benjamini-Hochberg FDR < 1%), including 156 up-regulated, and 111 down-regulated
304 proteins (Fig. 5a; and Supplementary Table 3). While no specific GO terms were enriched in
305 the set of down-regulated proteins, the inspection of the up-regulated proteins revealed a
306 strong enrichment in enzymes involved in the glucuronoxylan biosynthetic pathway (Fig. 5a,
307 b; and supplementary Table 3). Notably, the upregulated glucuronoxylan-related proteins
308 included glycosyltransferases required for the xylan backbone synthesis (IRX9/10/14), as well
309 as enzymes involved in its substitution (ESK1, GUX1, and GXM3)[2] (Fig. 5a, c). Visual
310 inspection of the RNA-seq aligned reads of the up-regulated glucuronoxylan-related genes
311 using the Integrative Genomics Viewer (IGV) and quantitative RT-PCR analysis of the
312 glucuronoxylan-related decoration genes confirmed that the observed proteomic changes in
313 msil2/4 were not associated with significant variations in mRNA levels (Supplementary Fig.
314 9a, b).

315 The outcomes of our histological and proteomic analyses suggest potential variations
316 in polysaccharide content and/or composition in the SCW of *msil2/4* mutant. To address this
317 point, we first performed glycome profiling on either untreated (-Tr) or H_2SO_4-treated (+Tr)
318 stem samples. The levels of glucose (Glc) and xylose (Xyl), which are the building blocks of
319 cellulose and xylan polysaccharides in the SCW were not substantially affected in the *msil2/4*
320 mutant (Fig. 5d). Similarly, no significant changes were observed in the levels of the primary
321 cell wall-related monosaccharides, galactose, rhamnose, arabinose, mannose, fucose and

322 galacturonic acid (Fig. 5d), indicating that the MSIL2/4 mutations did not significantly altered
323 cellulose or glucuronoxylan levels in the SCW.

324 To assess possible changes in glucuronoxylan decoration in *msil2/4* mutant, we
325 performed cytochemical analysis on Col-0 and *msil2/4* stem sections using two antixylan
326 antibodies (LM10 and LM11), whose binding specificities depend on the degree of
327 glucuronoxylan substitution[36]. LM11, which binds with similar affinity both unsubstituted and
328 substituted glucuronoxylan, shows a similar signal in the interfascicular and xylary fibers of
329 Col-0 and *msil2/4* mutant stems, confirming that the glucuronoxylan content was not
330 significantly modified in *msil2/4* compared to Col-0 (Fig. 5e). In contrast, LM10, which binds
331 less efficiently the substituted glucuronoxylan[36], shows a significant signal decrease in the
332 interfascicular fibers of *msil2/4* mutant (Fig. 5e; and Supplementary Table 4). Together, our
333 data indicate that the MSIL2/4 mutations impact the accumulation and activity of the
334 glucuronoxylan decoration machinery in the interfascicular fiber cells of Arabidopsis stem.

335

336 **MSILs restrain the degree of 4-O-methylation of glucuronoxylan in Arabidopsis.**

337 In Arabidopsis, the glucuronoxylan backbone is decorated predominantly by acetyl, GlcA,
338 and MeGlcA groups, that are deposited in a regular and controlled manner by a specific
339 enzymatic machinery[2,37,38]. Interestingly, the ESK1, GUX2, and GXM3 enzymes, which are
340 involved in the deposition of the acetyl and GlcA groups, as well as the methylation of GlcA
341 group, respectively, are over-accumulated in the *msil2/4* stem proteomics, providing a
342 possible explanation for the modification in glucuronoxylan decoration observed in the
343 interfascicular fiber cells of the *msil2/4* mutant. To investigate further the variations in
344 glucuronoxylan decoration incurred by the *msil2/4* mutation, we analyzed the patterns of
345 xylanase-released SCW oligosaccharide using matrix-assisted laser desorption ionization-
346 time-of-flight mass spectrometry (MALDI-TOF)[39]. Consistent with previously

347 observations[40,41], MALDI-TOF analysis of Col-0 glucurononoxylan oligosaccharides showed
348 the release of xylo-oligosaccharide pairs, evenly spaced by their degrees of polymerization
349 and acetylation, bearing a GlcA (m/z 1705.5/1747.5) or methylated GlcA (m/z 1719.5/1761.5)
350 substitution group (Fig. 6a). Interestingly, the level of GlcA branched xylo-oligomers was
351 strongly decreased in the *msil2/4* double mutant, and this whatever the complexity of the
352 oligomers considered (Fig. 6a; and Supplementary Fig. 10a). Importantly, the molecular
353 phenotypes observed in the *msil2/4* mutant were rescued in both *msil2/4+MSIL2*F1 and
354 *msil2/4+MSIL4*F1 complemented lines, confirming the specificity of the mutation (Fig. 6a). In
355 agreement with the MALDI–TOF outcome, quantification of the acidic sugar levels in
356 untreated (-Tr), or H2SO4-treated (+Tr) WT and *msil2/4* stem samples confirmed that the
357 release of GlcA moiety was strongly reduced in *msil2/4* double mutant, as opposed to the
358 galacturonic acid/GalA group located on pectins (Fig. 6b).

359 It has been previously proposed that the intermediate level of glucuronoxylan
360 methylation observed in Arabidopsis is due to an intrinsic rate-limiting activity of GXM
361 enzymes in vivo[42]. We noticed that in addition to GXM3, a second member of the GXM
362 family, GXM1[42], tends also to be over-expressed in *msil2/4* stem although at a fold change of
363 1.77 slightly below the fixed significance threshold in the proteomic analysis (Fig. 5a; and
364 Supplementary Table 3). This observations make us wonder whether the apparent decrease of
365 GlcA branched xylooligosaccharides in *msil2/4* could be due to an overmethylation of the
366 GlcA substituent. To address this question, we performed the xylanase digestion and MALDI-
367 TOF analysis in presence of an added spike-in oligosaccharide control (Pentaacetyl-
368 chitopentaose) for normalization of the MALDI_TOF data (Supplementary Fig. 10a). Upon
369 normalization, the MALDI-TOF data indicated that the decrease of the peaks corresponding
370 to the GlcA-branched xylooligosaccharides was correlated with a increase in corresponding
371 meGlcA branched xylooligosaccharides, (Fig. 6c; and Supplementary Fig. 10b; and

372 Supplementary Table 5), suggesting that the MSIL2/4 mutations unleash the activity of the
373 GXM1/3 in stem. RNA immunoprecipitation assays further confirmed that the *GXM3* and
374 *GXM1* mRNAs are efficiently pulled-down from MSIL4G-3, but not MSIL4G^{RRM}-3 plant
375 extracts (Fig. 6d), suggesting that the MSIL2/4 interaction network controls the 4-*O*-
376 methylation of glucuronoxylan by repressing the activity of the GXM1/3 mRNAs in stem.
377

378 **Discussion**

379 RNA-binding proteins (RBPs) are essential components of the gene regulation machinery that
380 govern the fate and expression of cellular RNA at post-transcriptional levels, including
381 processing, splicing, base modification, translation and degradation[13]. Sequence-based
382 bioinformatic, reverse and forward genetic, and affinity-based proteomic analyses have
383 converged to show that plants, like animals and fungi, harbor a large diversity of RBPs whose
384 functions in gene regulation and plant development are ill-defined[14]. In this study, we have
385 identified an RRM motif-containing RBP family, hereafter named as MSIL, whose members
386 share sequence, structural and functional similarities with the animal translational regulator
387 Musashi/MSI. We found that two of the four MSIL members, MSIL2 and MSIL4, function
388 redundantly to control various aspects of Arabidopsis development, including leaf senescence,
389 morphology of rosette leaves, and the rigidity of the inflorescence stem. We provide evidence
390 that the RRM domains of the MSIL2/4 proteins exhibit RNA-binding ability and are essential
391 for the activity of these proteins *in vivo*, an observation that is corroborated by the presence of
392 the MSIL2/4 proteins in an experimentally determined Arabidopsis mRNA-binding
393 proteome[14]. Further supporting a role for MSIL2/4 in mRNA-related transactions, affinity
394 purification-mass spectrometry analysis revealed that these proteins share a common network
395 of protein-protein interactions with well-known mRNA-binding proteins, including the
396 PAB2/4/8, RBP45/47, and the translation factor EIF4G. Notably, the animal polyA-binding

397 protein/PABP (homolog of the plant PABs) was previously shown to be a functional partner
398 of MSI that is essential for it activity in translation repression[18,26]. Interestingly, the MSIL2/4-
399 associated proteins also include the phylogenetically related RBGD2/4 proteins, two
400 constitutively expressed RBPs that are involved in heat resistance in Arabidopsis[20]. The
401 comparison of the MSIL2/4 and RBGD2/4 interactomes indicates only a partial overlap,
402 suggesting a certain level of subfunctionalization. In accordance with this observation, the
403 MSILs and RBGDs clearly diversify in function, since the *rbgd2/4* double mutant exhibits no
404 developmental phenotypes under normal growth conditions.

405 A penetrant developmental phenotype associated with the MSIL2/4 mutations is a loss
406 in stem rigidity that we could trace back to a defect in SCW formation, an important reservoir
407 of fixed carbon and a renewable resource of major environmental and economic importance[1].
408 Yet, with the exception of few microRNAs involved in the control of lignin polymer
409 synthesis[6,15,16], no information about RBPs-mediated posttranscriptional regulation of SCW
410 biosynthetic genes was previously reported. In this study, we show that MSIL2 and MSIL4,
411 function redundantly to promote SCW formation in the interfascicular fibers and that their
412 ability to bind RNA is essential to fulfill this function. Our conclusions are supported by
413 independent approaches which highlight changes of SCW thickness, lignin content, and xylan
414 decoration profile in the interfascicular fiber cells of the *msil2/4* mutant. The fiber-specific
415 effect of the MSIL2/4 mutations on SCW formation, is further supported by our evidence
416 showing that the *msil2/4* mutant grows normally and lacks the xylem collapses usually
417 observed in more pleiotropic SCW deficient mutants. Previous studies have shown that
418 specific master transcriptional regulators, such as NST1 and NST3, control cell-lineage
419 specific formation of SCW formation in Arabidopsis[5,31]. Our study extend this observation by
420 showing that RBP-mediated post-transcriptional mechanisms are also involved in cell-type
421 specific control of SCW formation in Arabidopsis. Notably, the presumed role of MSIL2/4 in

422 interfascicular fiber cells contrasts with the rather constitutive expression of the *MSIL2/4*
423 genes, suggesting that their activity depends on cell-context. This is in accordance with the
424 reported cell-type specific activities of ubiquitous animal MSI2 in cell renewal, cell
425 differentiation and cancer control[18,19]. Understanding how the activity of the *MSIL2/4* genes is
426 controlled by the cellular context will be essential from a basic point of view, providing a
427 unique model for studying RBP regulation in specific cell type in plants.

428 Our data indicate that the accumulation of gene transcripts of SCW biosynthesis
429 pathway enzymes and regulators are unaffected in *msil2/4*, suggesting that MSIL2/4 proteins
430 are likely to act at a translational level to regulate the SCW formation in interfascicular fiber
431 cells, an observation that is supported by the nature of the MSIL2/4 interactome. In this
432 regard, the stem proteomics gives us a glimpse into the changes in protein abundance that
433 could account for the observed *msil2/4* SCW phenotype. Indeed, our MS data reveal that the
434 glucuronoxylan biosynthesis machinery, including enzymes involved both in the synthesis
435 and substitution of the xylan backbone, is specifically over-accumulated in the *msil2/4* double
436 mutant. The molecular specificity of the observed glucuronoxylan phenotype with respect to
437 the changes in glucuronoxylan enzymatic machinery content raises several interesting
438 questions about the impact of the MSIL2/4 mutations on the SCW formation. First, we
439 noticed that the outcome of our proteomic data contrasts with the observed normal level of
440 glucuronoxylan deposition in the interfascicular fibers SCW, suggesting that independent
441 regulatory mechanisms act downstream of MSIL2/4 to control final glucuronoxylan content
442 in the SCW. One possible explanation for this observation could be that the activities of the
443 GUX, ESK, IRXs glucuronoxylan biosynthesis enzymes, in contrast to that of GXM3, are not
444 rate-limiting for glucuronoxylan synthesis *in vivo*. Alternatively, and in agreement with
445 studies that have reported that glucuronoxylan and cellulose contents are correlated *in vivo* in
446 xylan-deficient mutants[43], one could imagine that cellulose is the primary determinant of the

447 content in glucuronoxylan in the SCW, and that the glucuronoxylan produced in excess in the
448 golgi compartment of the *msil2/4* mutant is targeted for degradation upon secretion if
449 unbound to cellulose. In contrast, the concomitance between the glucuronoxylan
450 hypermethylation and the GXM3 overaccumulation phenotypes observed in *msil2/4* clearly
451 support the previous assumption that the GXM-mediated methylation is a rate-limiting step in
452 glucuronoxylan that can be overcome by GXM protein overexpression[42]. Our data support the
453 idea that MSIL2/4 play a specific role in the developmental control of the GXM protein
454 accumulation/activity, actively restraining the level of glucuronoxylan methylation in the
455 Arabidopsis interfascicular fiber cells. We currently posit that MSIL2/4 could control the
456 level of the GXM enzymes by directly regulating the translation of the *GXM* mRNAs or by
457 indirectly regulating the translation of an yet unknown translational regulator of the
458 glucuronoxylan biosynthesis pathway (XTRe) (Fig. 7). Future biochemical and functional
459 studies will be necessary to precisely understand the mechanism by which MSIL2/4 proteins
460 impact gene expression during SCW synthesis in Arabidopsis.

461 Our data indicate that, in addition to the reported change in glucuronoxylan
462 methylation, the MSIL2/4 mutations is also associated with a specific decrease of lignin
463 content in the SCW of the interfascicular fiber cells, whose origin remains unclear. On the
464 basis of the previous observations, we envision two possible mechanisms by which the
465 MSIL2/4 mutations could impact the lignin deposition in the fiber cells. In the first model, we
466 propose that MSIL2/4 positively control the lignin synthesis pathway in the interfascicular
467 cells by promoting the translation of lignin biosynthesis genes or that of a putative lignin
468 translational activator (LTAc) (Fig. 7). In this regard, although no GO enrichment could be
469 observed for lignin biosynthesis related genes in the proteomic analysis, two lignin
470 biosynthetic enzymes, CAD4 and PAL2, are significantly down-accumulated in *msil2/4*
471 mutant (Supplementary Table 3), an observation that could account for the observed decrease

472 in lignin content. However, previous published biochemical and genetic analyses do not
473 support this conclusion, as the single *cad4* and *pal2* knock-out mutants display no significant
474 alteration of lignin content due to genetic redundancy[9,44]. Moreover, thio-acidolysis analyses
475 showed that the ratio or degree of polymerization of the different lignin monomers remain
476 unchanged in the *msil2/4* mutant, reinforcing the idea that the lignin biosynthesis pathway is
477 functional in this mutant, leaving open the question of the specificity of MSIL2/4 action on
478 the lignin pathway (Fig.7). An alternative explanation for the decrease in lignin phenotype
479 observed in the *msil2/4* mutant would be that it results from an indirect cascade effect
480 primarily triggered by the *msil2/4*-dependent changes in glucuronoxylan methylation. In this
481 regard, it has been proposed that the 4-*O*-methylation of GlcA substituent results in the xylan
482 being more hydrophobic, possibly impacting its non-covalent interaction with lignin and the
483 capacity of lignin to assemble in the SCW[41,45,46]. Meanwhile, the hypermethylation of the
484 carbon-4 hydroxyl of the GlcA side chains could affect the capacity of glucuronoxylan to
485 form ether linkages to lignin (Fig. 7). Indeed, a study recently reported the existence of an
486 ether linkage between a carbon-6 hydroxyl of a mannosyl residue of a galactoglucomannan
487 hemicellulose and lignin in pine wood[47]. Although xylan has been proposed to bind
488 covalently to lignin via ether bonds involving xylose or Arabinose residues, it remains unclear
489 to date whether glucuronoxylan can form GlcA-dependent ether linkages with lignin, and the
490 functional relevance of such linkages remains to be demonstrated[48].
491

492 **Methods**

493 **Plant material.** *Arabidopsis thaliana* Col-0 and mutants *msil1-1* (*GABI_462A04*), *msil2-1*
494 (*Salk_066670*), *msil3-1* (*Salk_002477*) and *msil4-1* (*Salk_094167*) obtained from the SALK
495 Institute genomic analysis laboratory were used in this study. PAB2-RFP/PAB2R line has
496 been provided by Cécile Bousquet-Antonelli (LGDP, Université Perpignan). The primers

497 used to genotype lines are listed in Supplementary Table 4. Double mutants and quadruple
498 mutants were generated by crossing plants. Plants were either grown in soil or cultivated *in*
499 *vitro* on plates containing 2.20 g/l synthetic Murashige and Skoog (MS) (Duchefa) medium
500 and 0.5g/l MES, pH 5.7, and 7 g/l agar. To breakdown dormancy, seeds were incubated for 48
501 h at 6°C in the dark. Germination and development were performed in growth chambers, at
502 20°C, 60–75% hygrometry with a 16-h light/8-h dark photoperiod (100 µE m –2 s–1 light
503 [fluorescent bulbs with white 6500K spectrum, purchased from Sylvania]). For *in vitro*
504 culture, the seeds were surface-sterilized before been sown on plates, incubated for 48 h at
505 6°C in the dark, and placed in a growth cabinet at 20°C with a 16-h-d/8-h-dark cycle and 120
506 µE m –2 s–1 light (LEDs with white 4500K spectrum, purchased from Vegeled).
507
508 **Transgenic lines construction.** MSIL2F- and MSIL4F-tagged versions were produced from
509 Phusion-generated genomic DNA PCR products (promoter and coding region) using primers
510 TL976(SalI)-TL924(PstI) for MSIL2F and TL974(SalI)-TL975(BamHI) for MSIL4F. PCR
511 products were cloned in the CTL235 binary vector containing a Flag-HA tag in front of a
512 NosT terminator and the hygromycin resistance gene driven by a 35S promoter. To obtain the
513 MSIL2G and MSIL4G constructs used for fluorescence microscopy, the same PCR products
514 were cloned in the CTL579 binary vector in front of a EGFP tag. The obtained plasmids were
515 then used to transform *mgl2/4* plants via Agrobacterium transformation. MSIL4-
516 GFP/MSIL4G-tagged version used for complementation and RIP corresponds to the fusion of
517 a genomic PCR product containing *MSIL4* promoter (primers TL3527(HindIII)-TL3528(SalI))
518 fused with a second PCR cDNA fragment (primers TL3529(SalI)-TL3530(BamHI)) cloned
519 initially in pGEM T Easy (Promega). After sequencing the fusion DNA has been cloned in the
520 binary vector CTL579 containing GFP cDNA sequence. The MSIL4G-tagged construct was
521 finally introduced in *msil2/4* plants via Agrobacterium transformation. The rrm mutated

522 MSIL4GΔr version was produced as MSIL4G except that the cDNA fragment has been gene-
523 synthetized (Genecust) to modify RNA binding ability by having aspartic instead of
524 phenylalanine at position 9, 49 and 51 in RRM1 domain and in position 111, 151 and 153 in
525 RRM2 domain. Transgenic seeds selection was performed in the presence of 30 µg/l
526 hygromycin and the progenies were screened by PCR.
527
528 **Protein extraction and immunodetection**. Total plant extracts (up to 100 mg) were ground
529 in liquid nitrogen and proteins were extracted according to Hurkman and Tanaka[49] Before
530 migration, SDS–PAGE loading buffer was added and Coomassie staining was used to
531 calibrate loadings. Proteins were separated on SDS/ PAGE gels and blotted onto Immobilon-P
532 PVDF membrane (Cat. No. IPVH00010; MerckMillipore). Protein blot analysis was
533 performed using the Immobilon Western Chemiluminescent HRP Substrate (Cat. No.
534 WBKLS0500; MerckMillipore). Specific MSIL antibodies were raised in rabbits against
535 defined epitopes (Epitope MSIL1 (SCDGTSSTFGYNRIPS), MSIL2
536 (RLQEYFGKYGDLVE), MSIL3 (GYGVKPEVRYSPAVGN) and MSIL4
537 (TWRSPTPETEGPAPFS) (Eurogentec).
538
539 **Protein production and purification.** The MSIL4 RRM binding domains were amplified
540 using primers TL4128 and TL4129 either from MSIL4G construct for the wild type version or
541 from MSIL4GΔr for the mutated version. PCR products were then cloned in pET41 to obtain
542 His-tagged RRM domains. Tagged proteins were produced by induction in Escherichia coli
543 BL21 cells. Cultures of 200 mL of an ampicillin-resistant (100 mg mL–1) colony were grown
544 at 37°C and induced by 100mM isopropyl-b-D-galactopyranoside in the exponential phase
545 (optical density 0.5 at 500 nm). After induction, bacteria were harvested by centrifugation and
546 pellets were resuspended in 5 ml of binding buffer (His-bind buffer kit (Millipore), proteins

547 were extracted using a Constant cell disrupter system (Constant Systems) with a disruption
548 pressure of 2.35kbar and then purified using His-bind buffer kit (Millipore) following
549 manufacturer's recommendations.
550
551 **RNA binding assays.** Binding assays were performed using homopolymer RNA conjugated
552 to CNBr-activated Sepharose beads (Sigma **C9142)**. Either poly(A), poly(G), poly(C) or poly
553 (U) from Mercks (respectively P9403, P4903, P4404 and P9528) were used following
554 Sigma's recommendations. 0.7 to 1 mg of RNA were used for 200 mg sepharose beads. 0.45
555 mg of purified protein (MSIL4 RRM binding domain or mutated version) in 0.5 ml binding
556 buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 2.5 mM MgCl2, 2.5 % Triton X-100) were
557 incubated with 5 mg RNA conjugated Sepharose beads for 1 hour. After incubation, beads
558 were washed 5 times for 10 min each in 0.5 ml of binding buffer. Beads were then pelleted at
559 12,000 rpm, resuspended in SDS sample buffer, boiled for 20 min, and pelleted again.
560 Supernatants were transferred to fresh tubes and loaded on a 12% SDS-PAGE minigel for
561 electrophoresis (Bio-Rad) and protein blot analysis.
562
563 **RNA immunoprecipitation.** RIP was performed as described previously in Merret et al.[50]
564 using 400 mg of stem powder as starting material.
565
566 **Expression analysis**. Total RNA was isolated using the TRI re-agent (Cat. No. TR-118;
567 Euromedex) according to the manufacturer's instructions. Genomic DNA was then digested
568 out using the RQ1 RNase-Free DNase (Cat. No. M6101; Promega). cDNAs were obtained
569 from 400 ng of RQ1-treated RNA using 1 µl of GoScript Reverse Transcriptase (Cat. No.
570 A5003; Promega) in a 20-µl final volume reaction using random primers (Cat. No. C1181;
571 Promega) in the presence of 20 units of RNasin (Recombinant Ribonuclease Inhibitor, Cat.

572 No. C2511; Promega) and dNTP mix at a final concentration of 0.5 mM of each dNTP (Cat.
573 No. U1511; Promega). Semi-quantitative RT–PCR amplifications were performed on 1 µl of
574 cDNA in a 12.5-µl reaction volume to start with. The amplification of EF1αtranscripts was
575 used to equilibrate. qRT–PCR was performed on a Light Cycler 480 II machine (Roche
576 Diagnostics) by using the Takyon No ROX SYBR MasterMix blue dTTP kit (Cat. No. UF-
577 NSMT-B0701; Eurogentec). Each amplification reaction was set up in a 10-µl reaction
578 containing each primer at 300 nM and 1 µl of RT template in the case of total RNA with a
579 thermal profile of 95°C for 10 min and 40 amplification cycles of 95°C, 15 s; 60°C, 60 s.
580 Relative transcript accumulation was calculated using the ΔΔCt methodology[51], using
581 ACTIN2 as internal control. Average ΔΔCt represents three experimental replicates with
582 standard errors. PolyA+ RNAs were purified from RNA obtained by the TRI Reagent
583 extraction, using the PolyATtract mRNA Isolation System III (Cat. No. Z5310; Promega)
584 according to the manufacturer's instructions. cDNAs were then synthetized according to the
585 above protocol using 50 ng of polyA+ RNA as starting material. Semiquantitative and
586 quantitative RT–PCRs were performed as described above with a 1/25 dilution of cDNAs
587 from polyA+ RNA as template for qRT–PCR. The primers used for RT and qRT-PCR are
588 listed in Supplementary Table 6.
589
590 **RNA Sequencing.** Total RNA were extracted from the basal stem section of mature Col-0
591 and *msil2/4* mutant plants using the RNeasy Plant Mini Kit (Qiagen, Cat.#74904) and treated
592 onto the column with the RNase-Free DNase Set (Qiagen, Cat.#79254). Two replicates were
593 performed per plant lines. PolyA purification, library preparation using the Illumina TruSeq
594 stranded mRNA kit and library quality controls were performed by the BioEnvironment
595 Illumina Platform. The mRNA sequencing of Col-0 and msil2/4 mRNAs was done with
596 single-reads, 1x 75pb and 30–35x10^6 reads per library.

597

598 **RNASeq analysis.** For each library, 30–35x10^6 reads were obtained with 85–90% of the
599 bases displaying a Q-score ≥ 30, with a mean Q-score of 38 as assessed with Fastqc
600 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped against the
601 TAIR 10 genome using gtf annotation file and default parameters of TopHat2 v2.0.7[52].
602 Assembly and transcript quantification were performed with Cufflinks v2.2.1[53]. Finally, the
603 low expressed transcripts - less than 1 RPKM (reads per kilobase per million mapped reads)
604 in one of the libraries were filtered out. The differential analysis was conducted by Cuffdiff
605 software, belonging to the Trapnell suite. Foldchange (FC) was determined as the ratio
606 between the normalized read counts between mutant and wild-type.

607

608 **Fluorescence detection and tissue colorations.** GFP detections were performed on *in vitro*
609 grown one week seedling roots fixed in fixation buffer (50mM Pipes, 5mM EGTA, 5mM
610 MgSO4, pH7 added with 2% para-formaldehyde and 0.2% Triton). After 5 min vacuum
611 infiltration and 10 min agitation at room temperature, roots were washed twice with fixation
612 buffer and stored at 4°C until confocal observation. Confocal imaging was done using an
613 Zeiss Axio Observer Z1 LSM 700. Roots were observed in water and for GFP detection
614 samples were excited at 488 nm and the signal captured at 495–555 nm. For RFP detection,
615 excitation was performed at 555 nm and the signal captures at 600-700 nm. Images were
616 analysed with ImageJ. At least 15 samples per line were analyzed.

617 Stem coloration were performed from 5 cm length basal stem pieces conserved in
618 ethanol 80%. Stems were rehydrated first in ethanol 50% and later in water and then
619 embedded in agarose 5% before being transversally cut using a vibratome (Leica vt 1000s) to
620 obtain 120 micron thickness sections stored in water. The presence of lignin was determined
621 by staining with phloroglucinol for 30 seconds, giving a red coloration in the presence of

622 lignin cinnamaldehyde groups. Sections were observed using an inverted microscope (Leitz
623 DMRIBE, Leica Micro-systems, Wetzlar, Germany) and images were registered using a CCD
624 camera (Color Collview, Photonic Science, Milham, UK).

625 For cellulose detection, the brightener Calcofluor White ST [4,4'-bis(anilino-6-bis(2-
626 hydroxyethyl)-amino-s-triazine-ylamino)-2,2'-stilbene disulfonic acid; Cyanamide Co.,
627 Bound Brook, N.J] was used at a final concentration of 0.1%. The samples were incubated
628 with calcofluor for 3 min, washed three times with water before being observed. For
629 detection, excitation was performed at 405 nm and the signal captures at 434-496 nm. For
630 auto-fluorescence analysis stem sections were excited at 405 nm and the signal captures at
631 410-470 nm for blue. Pictures have been analyzed with LAS X (Leica Application Suite X).

632

633 **Biochemical analyses of SCWs.** Arabidopsis stems (5 cm sections at the bottom part of the
634 stems) were harvested from Col0 and *msil2/4* double mutant as four independent biological
635 replicates. Soluble extractives were eliminated as described by Ployet et al.[54]. Briefly, stems
636 were freeze dried for 48h, ground using a Mixer Mill MM 400 (Retsch), and extracted by hot
637 solvents (successively water, ethanol, ethanol/toluene (1:1 v/v) and acetone) to obtain
638 extractive-free stem residues (ESR). The determination of acetyl bromide (AcBr) lignin was
639 performed on 10mg of ESR as described by Ployet et al.[55]. All analyses were done in triplicate.
640 Fourier transform infrared spectroscopy (FT-IR) analysis was performed on 100-200 mg of
641 ESR. Spectra were recorded from ten technical replicates, in the range 400–4000 cm−1 with a
642 4 cm-1 resolution and 32 scans per spectrum using an attenuated total reflection (ATR)
643 Nicolet 6700 FT-IR spectrometer (Thermo Fisher, Illkirch-Graffenstaden, France) equipped
644 with a deuterated-triglycine sulfate (DTGS) detector. Spectra analyses were performed as
645 described in Dai et al.[56]. Saccharification yield was estimated with or without alkali
646 pretreatment, as described by Van Acker et al.[57]. Reducing sugar concentration was assessed

647 with dinitrosalicylic acid (DNS) reagent using 10μl of the supernatant after 9 h of reaction.
648 Enzyme activity was assessed at 0.25 filter paper units (FPU) ml^{-1} in our conditions.
649
650 **Mass spectrometry (MS)-based proteomic analyses.** Proteins from total inflorescence stem
651 extracts of three biological replicates of Col-0 and *msil2/4* mutant plants were solubilized in
652 Laemmli buffer and heated for 10 min at 95°C. They were then stacked in the top of a 4-12%
653 NuPAGE gel (Invitrogen), stained with Coomassie blue R-250 (Bio-Rad) before in-gel
654 digestion using modified trypsin (Promega, sequencing grade) as previously described[58]. The
655 resulting peptides were analyzed by online nanoliquid chromatography coupled to MS/MS
656 (Ultimate 3000 RSLCnano and Orbitrap Exploris 480, Thermo Fisher Scientific) using a 180-
657 min gradient. For this purpose, the peptides were sampled on a precolumn (300 µm x 5 mm
658 PepMap C18, Thermo Scientific) and separated in a 75 µm x 250 mm C18 column (Aurora
659 Generation 2, 1.6μm, IonOpticks). The MS and MS/MS data were acquired using Xcalibur
660 4.4 (Thermo Fisher Scientific).
661 Peptides and proteins were identified by Mascot (version 2.8.0, Matrix Science)
662 through concomitant searches against the Viridiplantae database (from Uniprot, Arabidopsis
663 thaliana (thale-cress) taxonomy, 136447 sequences) and a homemade database containing the
664 sequences of classical contaminant proteins found in proteomic analyses (human keratins,
665 trypsin..., 126 sequences). Trypsin/P was chosen as the enzyme and two missed cleavages
666 were allowed. Precursor and fragment mass error tolerances were set at respectively at 10 and
667 20 ppm. Peptide modifications allowed during the search were: Carbamidomethyl (C, fixed),
668 Acetyl (Protein N-term, variable) and Oxidation (M, variable). The Proline software[59]
669 (version 2.2.0) was used for the compilation, grouping, and filtering of the results
670 (conservation of rank 1 peptides, peptide length ≥ 6 amino acids, false discovery rate of
671 peptide-spectrum-match identifications < 1%[60], and minimum of one specific peptide per

672 identified protein group). MS data have been deposited to the ProteomeXchange Consortium
673 via the PRIDE partner repository[61]. Proline was then used to perform a MS1 label-free
674 quantification of the identified protein groups based on razor and specific peptides.
675 Statistical analysis was performed using the ProStaR software[62] based on the
676 quantitative data obtained with the four biological replicates analyzed per condition. Proteins
677 identified in the contaminant database, proteins identified by MS/MS in less than two
678 replicates of one condition, and proteins quantified in less than three replicates of one
679 condition were discarded. After log2 transformation, abundance values were normalized using
680 the variance stabilizing normalization (vsn) method, before missing value imputation (SLSA
681 algorithm for partially observed values in the condition and DetQuantile algorithm for totally
682 absent values in the condition). Statistical testing was conducted with limma, whereby
683 differentially expressed proteins were selected using a log2(Fold Change) cut-off of 1 and a p-
684 value cut-off of 0.00513, allowing to reach a false discovery rate close to 1% according to the
685 Benjamini-Hochberg estimator. Proteins found differentially abundant but identified by
686 MS/MS in less than two replicates, and detected in less than three replicates, in the condition
687 in which they were found to be more abundant were invalidated (p-value = 1).
688
689 **MALDI-TOF MS analyses of the patterns of xylanase-released SCW oligosaccharide.** 10
690 mg of dried alcohol insoluble cell wall residues were incubated with 10U.ml^{-1} of endo-1,4-D-
691 xylanase (E-XYLNP) from MEGAZYME previously desalted with centrifulgal filters
692 (Amicon Ultra-0.5ml) in 1 ml during 12hrs at 40°C. After boiling during 5 min to inactivate
693 enzyme activity, samples where filtred with 50KD centrifulgal filters (Amicon Ultra-0.5ml).
694 As positive control, 2 mg of wheat AX (MEGAZYMZE) were treated in a similar manner.
695 The samples were then analyzed by matrix-assisted laser desorption ionization-time of flight
696 (MALDI-TOF) MS using DMA/DHB (N,N-dimethylaniline/2,5-dihydroxybenzoic acid)

697 matrix[39]. The samples (1 µL) were deposited and then covered by the matrix (1 µL) on a
698 polished steel MALDI target plate. MALDI measurements were then performed on a rapifleX
699 MALDI-TOF spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a
700 Smartbeam 3D laser (355 nm, 10000 Hz) and controlled using the Flex Control 4.0 software
701 package. The mass spectrometer was operated with positive polarity in reflectron mode.
702 Spectra were acquired in the range of 560–3200 *m/z*. The detected species correspond to ions
703 in the form of sodium adducts.
704
705 **Thioacidolysis.** This method is adapted from Lapierre et al.[63]. Briefly, 10 mg of alcohol
706 insoluble cell wall residues were incubated in 3ml of dioxane with ethanethiol (10%), BF3
707 etherate (2.5%) containing 0.1% of heinecosane C21 diluted in CH2Cl2 at 100°C during 4
708 hours. 3ml of NaHCO3 (0.2 M) were added after cooling and mixed prior the addition of 0.1
709 mL of HCl (6 M). The tubes were wortexed after addition of 3 mL of dichloromethane and
710 the whole lower organic phase was collected in a new tube before concentration under
711 nitrogen atmosphere to approximately 0.5 ml. Then, 10 µL were trimethylsilylated (TMS)
712 with 100 µL of N,O-bis(trimethylsilyl) trifluoroacetamide and 10 µL of ACS-grade pyridine.
713 The trimethylsilylated samples were injected (1 µL) onto an Agilent 5973 Gas
714 Chromatography–Mass Spectrometry system. Specific ion chromatograms reconstructed at
715 *m/z* 239, 269 and 299 were used to quantify H, G and S lignin monomers respectively and
716 compared to the internal standard at m/z 57, 71, 85).
717
718 **Cell wall sugar analysis.** Neutral monosaccharide sugar content was determined by gas
719 chromatography after acid hydrolysis and conversion of monomers into alditol acetates as
720 described in Hoebler et al.[64] and Blakeney et al.[65] Gas Chromatography was performed on a

721 DB 225 capillary column (J&W Scientific, Folsorn, CA, USA; temperature 205 °C, carrier
722 gas H_2). Calibration was made with standard sugar solution and inositol as internal standard.
723 For acid sugar analysis, the uronic acid content was automatically quantified after
724 hydrolysis of polysaccharides in concentrated sulfuric acid (18 M) containing sodium
725 tetraborate (12.5 mM) or not, followed by m-hydroxydiphenyl colorimetric determination[66,67].
726 Galacturonic and Glucuronic acid were used as standard for quantification.
727

728 **Data availability**

729 The high-throughput sequencing data generated in this study are accessible at NCBI's Gene
730 Expression Omnibus (GEO) via GEO Series accession number GSE223732. The mass
731 spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via
732 the PRIDE partner repository with the dataset identifier PXD040207 (proteomic identification
733 of MSIL-interacting partners), and PXD040020 (proteomic analysis of total inflorescence
734 stem extracts).
735

736 **References**

737 1. Meents, M.J., Watanabe, Y. & Samuels, A.L. The cell biology of secondary cell wall
738 biosynthesis. *Ann Bot*. **121**, 1107-1125 (2018).
739 2. Scheller, H.V. & Ulvskov, P. Hemicelluloses. *Annu Rev Plant Biol*. **61**, 263-289 (2010).
740 3. Yokoyama, R. & Nishitani, K. Genomic basis for cell-wall diversity in plants. A
741 comparative approach to gene families in rice and Arabidopsis. *Plant cell Physiol*. **45**,
742 1111-1121 (2004).
743 4. Kumar, M., Campbell, L. & Turner, S. Secondary cell walls: biosynthesis and
744 manipulation. *J Exp Bot*. **67**, 515-531 (2016).
745 5. Taylor-Teeples, M. et al. An Arabidopsis gene regulatory network for secondary cell wal

746 synthesis. *Nature* **517**, 571-575 (2015).

747 6. Zhang, J., Xie, M., Tuskan, G.A., Muchero, W. & Chen, J.-G. Recent advances in the
748 transcriptional regulation of secondary cell wall biosynthesis in the woody plants. *Front*
749 *Plant Sci*. **9**, 1535 (2018).

750 7. Myburg, A.A. *et al*. The genome of Eucalyptus grandis. *Nature* **510**, 356-362 (2014).

751 8. Zhong, R. & Ye, Z.-H. Complexity of the transcriptional network controlling secondary
752 cell wall biosynthesis. *Plant Sci*. **229**, 193-207 (2014).

753 9. Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L. & Séguin, A.
754 CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes
755 involved in lignin biosynthesis in the floral stem of Arabidopsis. *Plant Cell* **17**, 2059-
756 2076 (2005).

757 10. Whitehead, C. et al. A glycosyl transferase family 43 protein involved in xylan
758 biosynthesis is associated with straw digestibility in *Brachypodium distachyon*. *New*
759 *Phytol*. **218**, 974-985 (2018).

760 11. Bonawitz, N.D. & Chapple, C. Can genetic engineering of lignin deposition be
761 accomplished without an unacceptable yield penalty? *Curr Opin Biotechnol*. **24**, 336-343
762 (2013).

763 12. Wang, Y. et al. Genetic modification of plant cell walls to enhance biomass yield and
764 biofuel production in bioenergy crops. *Biotechnol Adv*. **34**, 997-1017 (2016).

765 13. Singh, G., Pratt, G., Yeo, G.W. & Moore, M.J. The clothes make the mRNA: past and
766 present trends in mRNP fashion. *Ann Rev Biochem*. **84**, 325-354 (2015).

767 14. Reichel, M. et al. In planta determination of the mRNA-binding proteome of Arabidopsis
768 etiolated seedlings. *Plant Cell* **28**, 2435-2452 (2016).

769 15. Wang, C.Y. et al. MiR397b regulates both lignin content and seed number in Arabidopsis
770 via modulating a laccase involved in lignin biosynthesis. *Plant Biotechnol J*. **12**, 1132-

771 1142 (2014).

772 16. Zhao, Y. et al. MicroRNA857 is involved in the regulation of secondary growth of
773 vascular tissues in Arabidopsis. *Plant physiol*. **169**, 2539-2552 (2015).

774 17. Chai, G. et al. Arabidopsis C3H14 and C3H15 have overlapping roles in the regulation of
775 secondary wall thickening and anther development. *J Exp Bot*. **66**, 2595-2609 (2015).

776 18. Fox, R.G., Park, F.D., Koechlein, C.S., Kritzik, M. & Reya, T. Musashi signaling in stem
777 cells and cancers. *Annu Rev Cell Dev Biol*. **31**, 249-267 (2015).

778 19. Kudinov, A.E., Karanicolas, J., Golemis, E.A. & Boumber, Y. Musashi RNA-binding
779 proteins as cancer drivers and novel therapeutic targets. *Clin Cancer Res*. **23**, 2143-2153
780 (2017).

781 20. Zhu, S. & al. Liquid-liquid phase separation of RBGD2/4 is required for heat stress
782 resistance in Arabidopsis. *Dev Cell*. **57**, 583-597 (2022).

783 21. Chavali, P.L. et al. Neurodevelopmental protein Musashi-1 interacts with the Zika
784 genome and promotes viral replication. *Science* **357**, 83-88 (2017).

785 22. Nguyen, D.T.T. et al. HyperTRIBE uncovers increased MUSASHI-2 RNA binding
786 activity and differential regulation in leukemic stem cells. *Nat Commun*. **11**, 2026 (2020).

787 23. Biasani, M. et al. SWISS-MODEL: modelling protein tertiary and quarternary structure
788 using evolutionary information. *Nucleic Acids Res*. **42**, W252-258 (2014).

789 24. Daubner, G.M., Cléry, A. & Allain, F.H. RRM-RNA recognition: NMR or
790 crystallography… and new findings. *Curr Opin Struct Biol*. **23**,100-108 (2013).

791 25. Weill, L., Belloc, E., Castellazzi, C.L. & Mendez, R. Musashi 1 regulates the timing and
792 extent of meiotic mRNA translational activation by promoting the use of specific CPEs.
793 *Nat Struct Mol Biol*. **24**, 672-681 (2017).

794 26. Cragle, C.E. et al. Musashi interaction with poly(A)-binding protein is required for
795 activation of target mRNA translation. *J Biol Chem*. **294**, 10969-10986 (2019).

796 27. Scheer, H. et al. The TUTase URT1 connects decapping activators and prevents the
797 accumulation of excessively deadenylated mRNAs to avoid siRNA biogenesis. *Nat*
798 *Commun*. **12**, 1298 (2021).
799 28. Chantarachot, T. & Bailey-Serres, J. Polysomes, stress granules, and processing bodies: a
800 dynamic triumvirate controlling cytoplasmic mRNA fate and function. *Plant Physiol*. **176**,
801 254-269 (2018).
802 29. Scutenaire, J. et al. The YTH domain protein ECT2 is an m^6A reader required for normal
803 trichome branching in Arabidopsis. *Plant Cell* **30**: 986-1005 (2018).
804 30. Mahboubi, H. & Stochaj, U. Cytoplasmic stress granules: Dynamic modulators of cell
805 signaling and disease. *Biochim Biophys Acta Mol Basis Dis*. **1863**, 884-895 (2017).
806 31. Mitsuda, N. et al. NAC transcription factors, NST1 and NST3, are key regulators of the
807 formation of secondary walls in Woody tissues of Arabidopsis. *Plant Cell* **19**, 270-280
808 (2007).
809 32. Shi, D. et al. Tissue-specific transcriptome profiling of the Arabidopsis inflorescence
810 stem reveals local cellular signatures. *Plant Cell* **33**, 200-223 (2021).
811 33. Herth, W. & Schnepf, E. The fluorochrome, calcofluor white, binds oriented to structural
812 polysaccharide fibrils. *Protoplasma* **105**, 129-133 (1980).
813 34. Alonso-Simón, A. et al. The use of FTIR spectroscopy to monitor modifications in plant
814 cell wall architecture caused by cellulose biosynthesis inhibitors. *Plant Signal Behav*. **6**,
815 1104-1110 (2011).
816 35. Miedes, E., Vanholme, R., Boerjan, W. & Molina, A. The role of the secondary cell wall
817 in plant resistance to pathogens. *Front Plant Sci*. **5**, 358 (2014).
818 36. McCartney, L. et al. Differential recognition of plant cell walls by microbial xylan-
819 specific carbohydrate-binding modules. *Proc Natl Acad Sci USA*. **103**, 4765-4770 (2006).
820 37. Grantham, N.J. et al. An even pattern of xylan substitution is critical for interaction with

821 cellulose in plant cell walls. *Nat Plants* **11**, 859-865 (2017).

822 38. Wierzbicki, M.P., Maloney, V., Mizrachi, E. & Myburg, A.A. Xylan in the middle:
823 understanding xylan biosynthesis and its metabolic dependencies toward improving wood
824 fiber for industrial processing. *Front Plant Sci*. **25**, 176 (2019).

825 39. Ropartz, D et al. Performance evaluation on a wide set of matrix-assisted laser desorption
826 ionization matrices for the detection of oligosaccharides in a high-throughput mass
827 spectrometric screening of carbohydrate depolymerizing enzymes. *Rapid Commun Mass*
828 *Spectrom*. **25**, 2059-2070 (2011).

829 40. Mortimer, J. C. et al. Absence of branches from xylan in Arabidopsis *gux* mutants reveals
830 potential for simplification of lignocellulosic biomass. *Proc Natl Acad Sci USA*. **107**,
831 17409-17414 (2010).

832 41. Urbanowicz, B.R. et al. 4-O-methylation of glucuronic acid in Arabidopsis
833 glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proc*
834 *Natl Acad Sci USA*. **109**, 14253-14258 (2012).

835 42. Yuan, Y., Teng, Q., Lee, C., Zhong, R. & Ye, Z.-H. Modification of the degree of 4-O-
836 methylation of secondary cell wall glucuronoxylan. *Plant Sci*. **219-220**, 42-50 (2014).

837 43. Wu, A.-M. et al. The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are
838 critical for glucuronoxylan biosynthesis during secondary cell wall formation. *Plant J*. **57**,
839 718-731 (2009).

840 44. Rohde, A. et al. Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis
841 thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and
842 carbohydrate metabolism. Plant cell 16, 2749-2771 (2004).

843 45. Busse-Wicher, M. et al. Evolution of xylan substitution patterns in gymnosperms and
844 angiosperms: implications for xylan interaction with cellulose. *Plant Physiol*. **171**, 2418-
845 2431 (2016).

846 46. Kang, X. et al. Lignin-polysaccharide interactions in plant secondary cell walls revealed
847 by solid-state NMR. *Nat Commun*. **10**, 347 (2019).
848 47. Nishimura, H., Kamiya, A., Nagata, T., Katahira, M. & Watanabe, T. Direct evidence for
849 α ether linkage between lignin and carbohydrates in wood cell walls. *Sci Rep*. **8**, 6538
850 (2018).
851 48. Terret, O.M. and Dupree, P. Covalent interactions between lignin and hemicelluloses in
852 plant secondary cell walls. *Curr Opin Biotechnol*. **56**, 97-104 (2019).
853 49. Hurkman, W.J. & Tanaka, C.K. Solubilization of plant membrane proteins for analysis by
854 two-dimensional gel electrophoresis. Plant Physiol. 81, 802-806 (1986).
855 50. Merret, R. et al. Heat shoxk protein HSP101 affects the release of ribosomal protein
856 mRNAs for recovery after heat shock. *Plant Physiol*. **174**, 1216-1225 (2017).
857 51. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time
858 quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**,402-408 (2001).
859 52. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
860 **9**, 357-359 (2012).
861 53. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of
862 insertions, deletions and gene fusions. *Genome Biol*. **14:** R36 (2013).
863 54. Ployet, R. et al. Long cold exposure induces transcriptional and biochemical remodelling
864 of xylem secondary cell wall in Eucalyptus. *Tree Physiol*. **38**, 409-422 (2018).
865 55. Ployet, R. et al. A systems biology view of wood formation in Eucalyptus grandis trees
866 submitted to different potassium and water regimes. *New Phytol*. 223, 766-782 (2019).
867 56. Dai, Y. et al. Implementing the CRISPR/Cas9 technology in Eucalyptus hairy roots uisng
868 wood-related genes. *Int J Mol Sci*. **21**, 3408 (2020).
869 57. Van Acker, R., Vanholme, R. & Piens K. Saccharification protocol for small-scale
870 lignocellulosic biomass samples to test processing of cellulose into glucose. *Bio-Protocol*.

871 **6**, 1701 (2016).

872 58. Casabona, M.G., Vandenbrouck, Y., Atree, I. & Couté, Y. Proteomic characterization of
873 Pseudomonas aeruginosa PAO1 inner membrane. *Proteomics* **13**, 2419-2423 (2013).

874 59. Bouyssié, D. et al. Proline: an efficient and user-friendly software suite for large-scale
875 proteomics. *Bioinformatics* **36**, 3148-3155 (2020).

876 60. Couté, Y., Bruley, C. & Burger, T. Beyond target-decoy competition: stable validation of
877 peptide and protein identifications in mass spectrometry-based discovery proteomics.
878 *Anal Chem*. **92**, 14898-14906 (2020).

879 61. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019:
880 improving support for quantification data. *Nucleic Acids Res*. **47**, D442-D450 (2019).

881 62. Wieczorek, S. et al. DAPAR & ProStaR: software to perform statitical analyses in
882 quantitative discovery proteomics. *Bioinformatics* **33**, 135-136 (2017).

883 63. Lapierre, C., Pollet, B. & Rolando, C. New insights into the molecular architecture of
884 hardwood lignins by chemical degradative methods. *Res Chem Intermed*. **21**, 397-412
885 (1995).

886 64. Hoebler C, Barry J.-L., David A. & Delort-Laval, J. Rapid acid hydrolysis of plant cell
887 wall polysaccharides and simplified quantitative determination of their neutral
888 monosaccharides by gas-liquid chromatography. *J Agric Food Chem*. **37**, 360–365 (1989).

889 65. Blakeney, A.B., Harris, P.J., Henry, R.J. & Stone, B.A. A simple and rapid preparation of
890 alditol acetates for monosaccharide analysis. *Carbohydr Res*. **113**, 291–299 (1983).

891 66. Blumenkrantz, N. & Asboe-Hansen, G. New method for quantitative determination of
892 uronic acids. *Analytical Biochemistry* **54**, 484-489 (1973).

893 67. Thibault, J.F. Automated-method for the determination of pectic substances.
894 *Lebensmittel-Wissenschaft Und-Technologie* **12**, 247-251 (1979).

895

896 **Acknowledgements**

897 We thank Aurélie Le Ru for the histological staining and cell wall thickness measurements,
898 Hua Cassan-Wang for helping to the preparation of stem sections for staining, and Michele
899 Laudié for preparation of the mRNA libraries. This work was supported by the Centre
900 National de la Recherche Scientific (CNRS), and grants ANR-08-BLAN-0206 and 12-BSV6-
901 0010 from Agence National de la Recherche (ANR) to TL. This study was supported by the
902 "Ecole Universitaire de Recherche (EUR) TULIP-GS (ANR-18-EURE-0019), and the
903 EPIPLANT Groupement de Recherche (CNRS, France). The proteomic experiments
904 performed by LB and YC were partially supported by Agence Nationale de la Recherche
905 under projects ProFI (Proteomics French Infrastructure, ANR-10-INBS-08) and GRAL, a
906 program from the Chemistry Biology Health (CBH) Graduate School of University Grenoble
907 Alpes (ANR-17-EURE-0003). The funders had no role in study design, data collection and
908 analysis, decision to publish or preparation of the paper.

909

910 **Author contributions**

911 Study design: T.L., F.M., R.S., M.F., DP., and N.B.E.; mutant characterization and molecular
912 biology: A.K., D.P., C.P., R.M., and J.A.; RNA binding analysis: N.B.E.; MSIL co-
913 purification experiments, mass spectrometry and statistical analysis of co-IP data: D.G. and
914 P.H.; Total inflorescence stem proteomic and statistical analysis: L.B. and Y.C.;
915 Polysaccharide biochemical analysis: L.L-B. and R.S.; Lignin biochemical and FTIR analysis:
916 F.M.; immunohistochemical staining: Y.M.; MALDI-TOF analysis and statistical analysis:
917 R.S. and M.F.; writing: T.L., F.M., R.S., M.F., D.P., and N.B.E.; visualisation: T.L.;
918 supervision: T.L. and N.B.E.

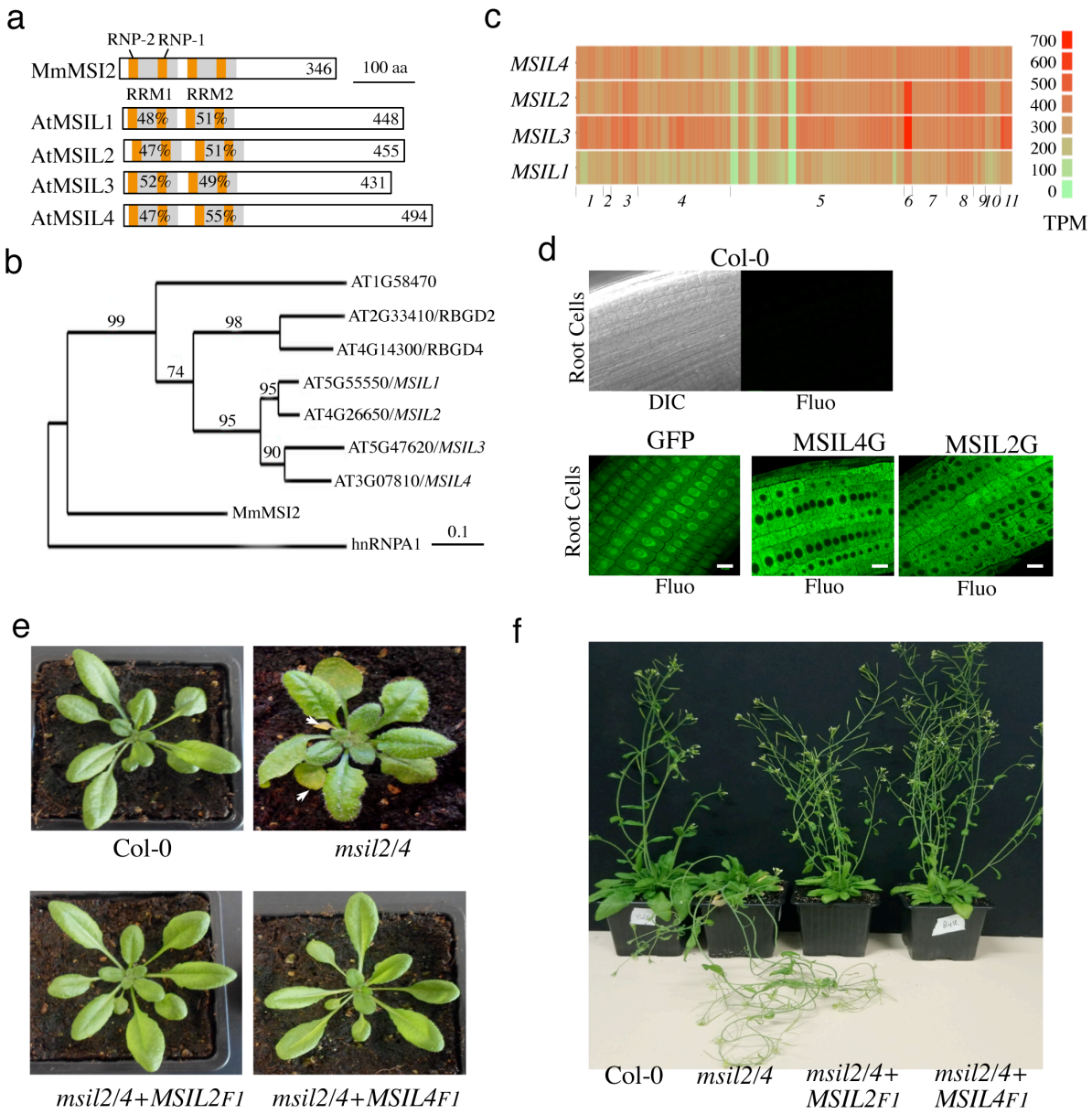
919

920 **Competing interests**

921 The authors declare no competing interests.

922

923 **Figures**



924

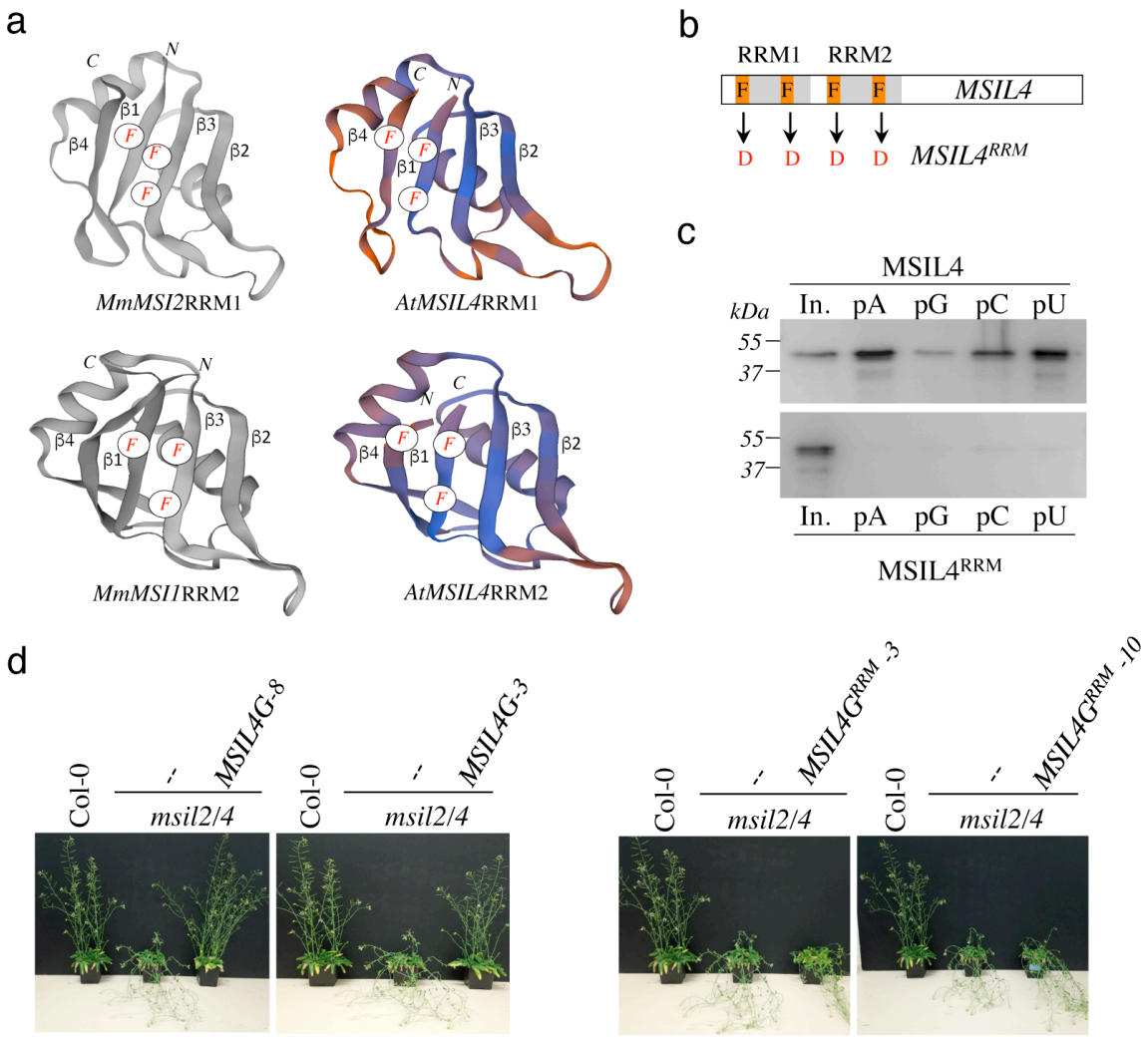
925 **Fig. 1 | Musashi-like MSIL2/4 proteins redundantly control development in Arabidopsis.**

926 a) Schematic representation of the Musashi-like (MSIL) protein family in *Arabidopsis*

927 *thaliana*. The invariant RNP1 and RNP2 motifs within the conserved RRM domains are

928 indicated. Numbers refer to amino acid identities between the Mouse Musashi/MSI RRM1/2

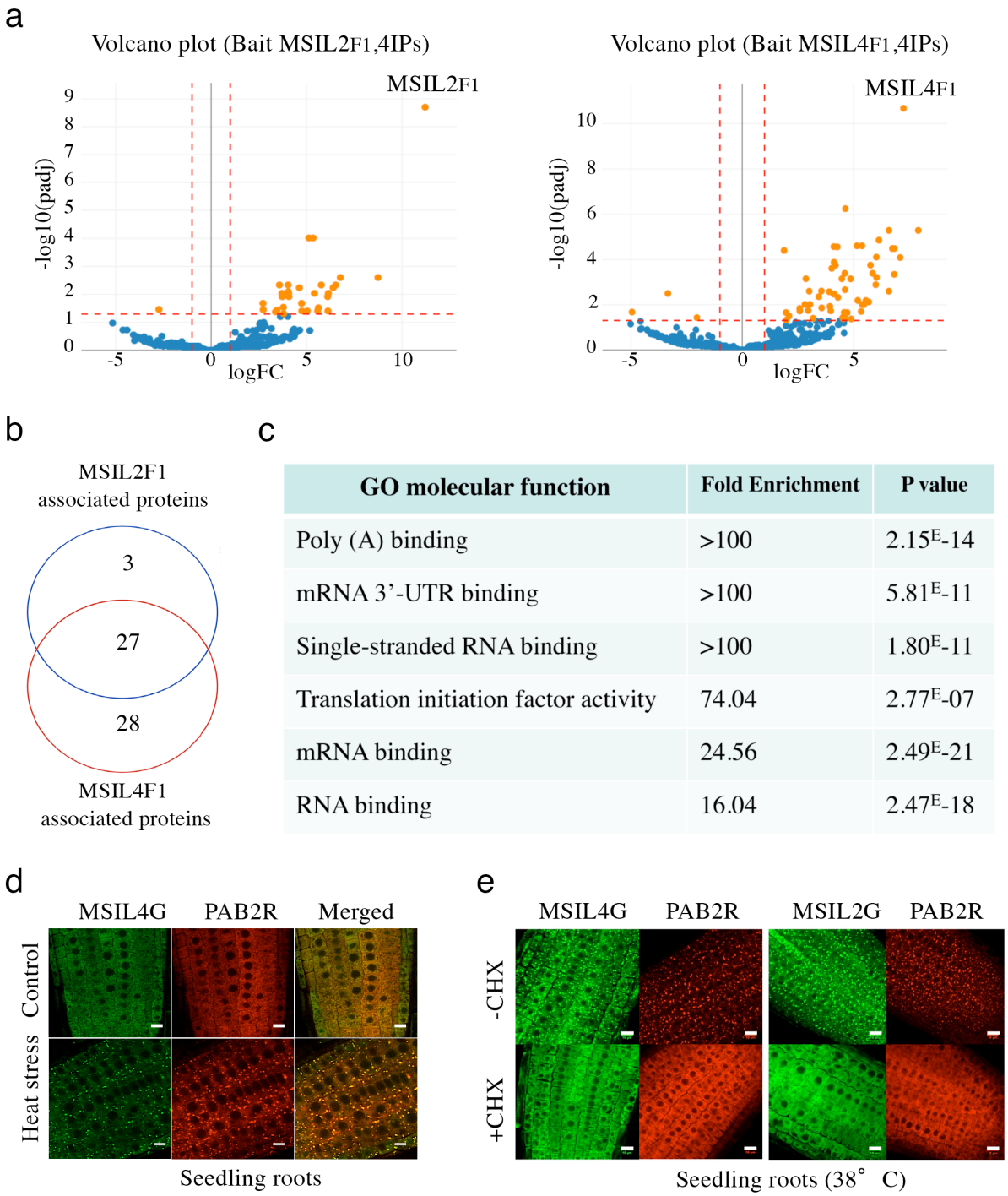
929 domains and the corresponding domains in Arabidopsis MSIL homologs. b) Evolutionary
930 relationships between MSILs and related RNA-binding proteins. The scale bar indicates the
931 rate of evolutionary change expressed as number of amino acid substitutions per site. c) RNA-
932 seq expression map of *MSIL* genes extracted from ARAPORT11. Shading is a $\log_2$ scale of
933 transcripts per million (TPM). d) General overview of the roots of 6-days-old wild type
934 Arabidopsis seedlings (Col-0) or transgenic seedlings expressing either a free GFP protein or
935 GFP-tagged versions of MSIL2/MSIL2G and MSIL4/MSIL4G. Scale bar, 10 μm. e)
936 Photographs of representative rosettes of Col-0, *msil2/4*, *msil2/4-MSIL2*F1 and *msil2/4-*
937 *MSIL4*F1 plants. f) Photographs of representative inflorescence stems of Col-0, *msil2/4*,
938 *msil2/4-MSIL2*F1 and *msil2/4-MSIL4*F1 plants. Abbreviations: Mus musculus Musashi2
939 (MmMsi2); human heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1).
940
941
942
943
944
945
946
947
948
949
950
951



952

953 **Fig. 2 | RRM-dependent RNA-binding activity is essential for MSIL function *in planta*.**
954 Experimentally determined structures of the RRM1 domain of MmMSI2 (PDB ID: 6C8U)
955 and the RRM2 domain of MmMSI1 (PDB ID: 5x3y), and models of the RRM domains of
956 MSIL4 generated using the homology-modeling server SWISS-MODEL. The α-helices and
957 β-sheets of RRM domain as well as the phenylalanine residues that contact RNA bases are
958 highlighted. b) Schematic representation of the mutations introduced in the RRM domains of
959 MSIL4. c) Binding assays of MSIL4 RRM domains on ssRNA homopolymers *in* vitro.
960 Single-stranded polyA (pA), polyG (pG), polyC (pC) or polyU (pU) RNAs were subjected to
961 binding with the His-tagged recombinant RRM domains from MSIL4 (upper part) or its
962 mutated version ($MSIL4^{RRM}$, lower part). An anti-His antibody was used for detection. In.

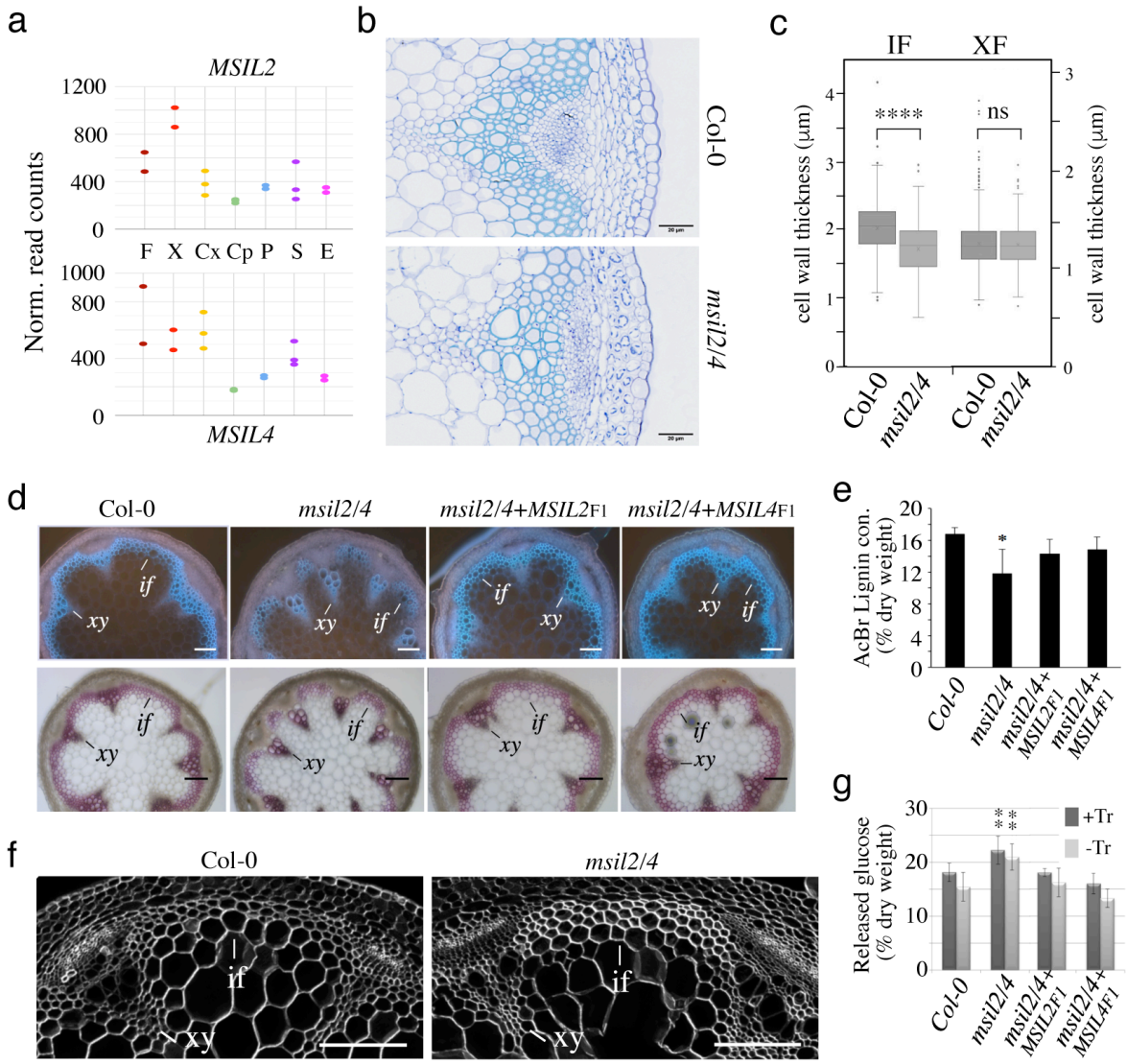
963 represents the input fraction. d) RRM-dependent RNA-binding activity is essential for
964 MSIL2/4 function in stem. Photographs of representative inflorescence stems of Col-0 and
965 *msil2/4* mutants complemented or not with a WT version of MSIL4 (MSIL4G-8 and 3) or an
966 rrm mutant (MSIL4G^{RRM}-3 and 10).
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984



GO molecular function	Fold Enrichment	P value
Poly (A) binding	>100	2.15^{E}-14
mRNA 3'-UTR binding	>100	5.81^{E}-11
Single-stranded RNA binding	>100	1.80^{E}-11
Translation initiation factor activity	74.04	2.77^{E}-07
mRNA binding	24.56	2.49^{E}-21
RNA binding	16.04	2.47^{E}-18

Fig. 3| MSIL2/4 protein interactomes are enriched in proteins involved in 3'-UTR binding and translation regulation. a) The semi-volcano plots show the enrichment of proteins co-purified with MSIL2F1 and MSIL4F1 as compared with control IPs. Y- and X-axis display adjusted p-values and fold changes, respectively. The dashed lines indicate the threshold above which proteins are significantly enriched/depleted (fold change > 2; adjP <

991 0.05). b) Venn diagram showing the overlap between the MSIL2 and the MSIL4
992 interactomes. c) PANTHER-based classification of the GO molecular function that are
993 overrepresented among the MSIL-interacting proteins. d) Confocal monitoring of the
994 colocalization of MSIL2G, MSIL4G and PAB2R in root tips of 8-d-old seedlings were
995 monitored after 30 min of exposure to 38°C (heat stress) or 20°C for control treatment. e)
996 Cycloheximide treatment inhibits the formation of MSILG and PAB2R cytosolic foci. The
997 transgenic plants expressing the stress granule marker PAB2R was used as a control. Root of
998 7-day-old seedlings expressing either MSIL4G, MSIL2G, or PAB2R were monitored after 1
999 hour of exposure to 38°C in the presence of DMSO control treatment (DMSO) or in the
1000 presence of cycloheximide inhibitor (+CHX). Scale bar, 10 μm.
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012



1013

1014 **Fig. 4 | MSIL proteins regulate the molecular architecture of SCW in Arabidopsis fibers.**

1015 a) Expression profiles of the *MSIL2* and *MSIL4* genes in the inflorescence stem tissues as
1016 retrieved from the Arabidopsis inflorescence stem tissue-specific transcriptome database[40]
1017 (https://arabidopsis-stem.cos.uni-heidelberg.de/). The results of two replicates are shown. F,
1018 fibers; X, xylem vessels; Cx Cambium (xylem side); Cp, Cambium (Phloem side); P, Phloem;
1019 S, Starch sheath; E, Epidermis. b) Cross-sections of Col-0 and *msil2/4* stems showing
1020 vascular bundle and interfascicular fiber cells stained with Toluidine blue. c) Measurements
1021 of the SCW thickness of xylary fiber and interfascicular fiber cells. Values are means
1022 (n>400)± SEM. Data were analyzed by unpaired Student's test. Asterisks indicate significant

1023 differences relative to Col-0; ****p<0.0001. d) Top: lignin autofluorescence under ultraviolet
1024 (UV) using confocal microscopy in Col-0, *msil2/4* mutant, and complemented *msil2/4-*
1025 *MSIL2*F1 and *msil2/4-MSIL4*F1 plant stem sections. xy: xylem fibers; if: interfascicular fibers.
1026 Bottom: Phloroglucinol-HCl staining of Col-0, *msil2/4*, and complemented *msil2/4-MSIL2*F1
1027 and *msil2/4-MSIL4*F1 plant stem sections. Scale bar, 100 µm. e) Histograms represent acetyl
1028 bromide lignin content, expressed as % of dry weight. Lignin content was measured in the
1029 bottom section of mature stems (3 biological replicates per line). Significant differences
1030 between Col-0, *msil2/4+MSIL2*F1, *msil2/4+MSIL4*F1, and *msil2/4* plant lines are indicated by
1031 asterisks*, *P-value*< 0.01 according to Student's t test (n = 6 to 10). f) Confocal microscopy
1032 of Calcofluor-white staining of cross sections of Col-0 and *msil2/4* mutant inflorescence
1033 stems. Scale bar, 100 µm. g) Glucose yield after incubation of Col-0 and *msil2/4*
1034 lignocellulosic material with cellulases either without pre-treatment (-Tr) or after sodium
1035 hydroxide pre-treatment (+Tr).

1036

1037

1038

1039

1040

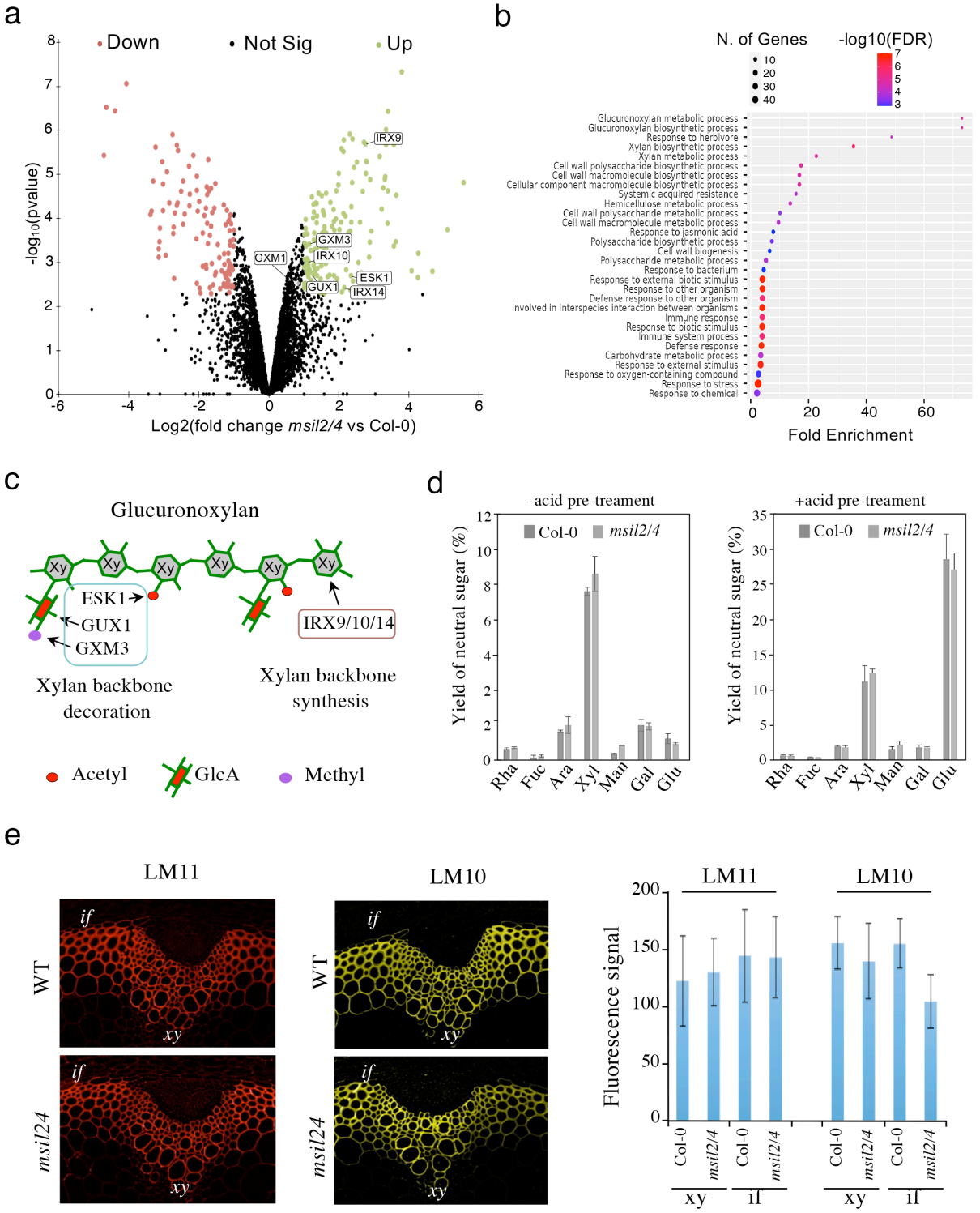
1041

1042

1043

1044

1045



1046

1047 **Fig. 5 | The accumulation and activity of the glucuronoxylan decoration machinery are**

1048 **altered in *msil2/4* mutant.**

1049 a) MS-based quantitative comparison of Col-0 and *msil2/4* inflorescence stem proteomes.

1050 Volcano plot displaying the differential abundance of proteins in Col-0 and *msil2/4* cells

1051 analyzed by MS-based label-free quantitative proteomics. The volcano plot represents the -
1052 log10 (limma p-value) on y axis plotted against the log2 (Fold Change msil2/4 vs Col-0) on x
1053 axis for each quantified protein. Green and red dots represent proteins found significantly
1054 enriched respectively in *msil2/4* and Col-0 Arabidopsis stems (log2(Fold Change) ≥ 1 and -
1055 log10(p-value) ≥ 2.29, leading to a Benjamini-Hochberg FDR = 1.01 %). The up-regulated
1056 proteins involved the glucuronoxylan biosynthetic pathways are indicated. b) Gene Ontology
1057 (GO) analysis of the proteins that are significantly up-regulated in the *msil2/4* mutant. The
1058 gene ontology analysis of DEGs was performed using ShinyGO v0.76 software. Lollipop
1059 diagrams provide information about GO fold enrichment, significance (FDR in log10), and
1060 number of genes in each pathway. c) Schematic model of glucuronoxylan substitution
1061 patterns in Arabidopsis and the enzymes involved. IRX9/10/14, glycosyltransferases involved
1062 in the synthesis of xylan (Xy) backbone. ESK1, eskimo1. GUX1, glucuronic acid substitution
1063 of xylan1; GXM3, glucuronoxylan methyltransferase3 are involved in the glucuronoxylan
1064 decoration. d) Neutral monosaccharide composition of alcohol insoluble residue (AIR)
1065 extracted from inflorescence stems of wild-type and *msil2/4* mutant plants that have been pre-
1066 hydrolyzed or not with acid. Rha, rhamnose; Fuc, fucose; Ara, arabinose; Xyl, xylose; Man,
1067 mannose; Gal, galactose; Glc, glucose. e) Left: Immunofluorescence labeling of transverse
1068 sections of Col-0 and *msil2/4* stems with the LM10 and LM11 antixylan antibodies. Xy,
1069 xylem fibers; if, interfascicular fibers. Right: Quantification of the fluorescence was done
1070 using ImageJ software and processed according to Supplementary table 4.

1071

1072

1073

1074

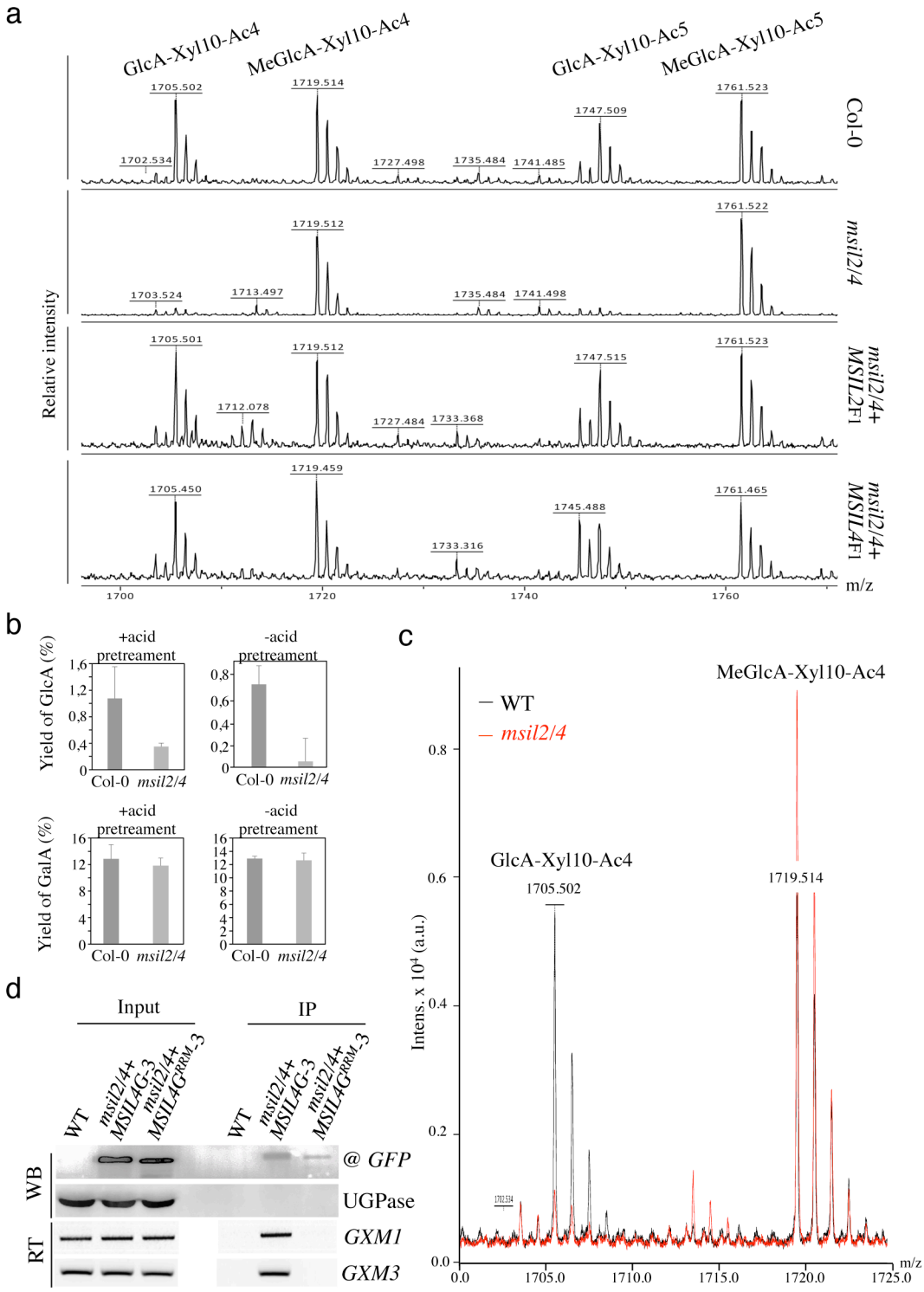


Fig. 6 | MSILs restrain the degree of 4-O-methylation of glucuroxylan in Arabidopsis inflorescence stem. a) MALDI-TOF mass spectra of xylooligosaccharides generated by xylanase digestion of xylan from Col-0, *msil2/4*, and complemented *msil2/4-MSIL2*F1 and *msil2/4-MSIL4*F1 inflorescence stem materials. The ions at m/z 1705/1747 and 1719/1761

1080 correspond to acetylated xylo-decapolysaccharides bearing a GlcA residu (GlcA-Xyl10-
1081 Ac4//GlcA-Xyl10-Ac5) or a methylated GlcA residue (MeGlcA-Xyl10-Ac4//MeGlca-Xyl10-
1082 Ac5). b) Acidic monosaccharide composition of alcohol insoluble residue (AIR) extracted
1083 from inflorescence stems of wild-type and *msil2/4* mutant plants that have been pre-
1084 hydrolyzed or not with acid. GalA, galacturonic acid; GlcA, glucuronic acid. c) Relative
1085 changes in unmethylated/methylated GlcA decapolysaccharide ratio in Col-0 (black) and
1086 *msil2/4* mutant (red) as controlled by the addition of external spike-in control corresponding
1087 to a pentaacetyl-chitopentaose. d) GFP-based RNA-IP assays. Top panel (WB): western blots
1088 performed using antiGFP or antiUGPase antibodies on protein fractions from inputs or
1089 antiGFP immunoprecipitates from WT and the *msil2/4* mutant plants expressing the WT
1090 (MSIL4G-3) or RRM mutant (MSIL4G^{RRM}-3) MSIL4-GFP fusions. Bottom panel (RT) :
1091 Corresponding RT-PCR using *GXM1* or *GXM3* specific primers on RNA fractions.

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

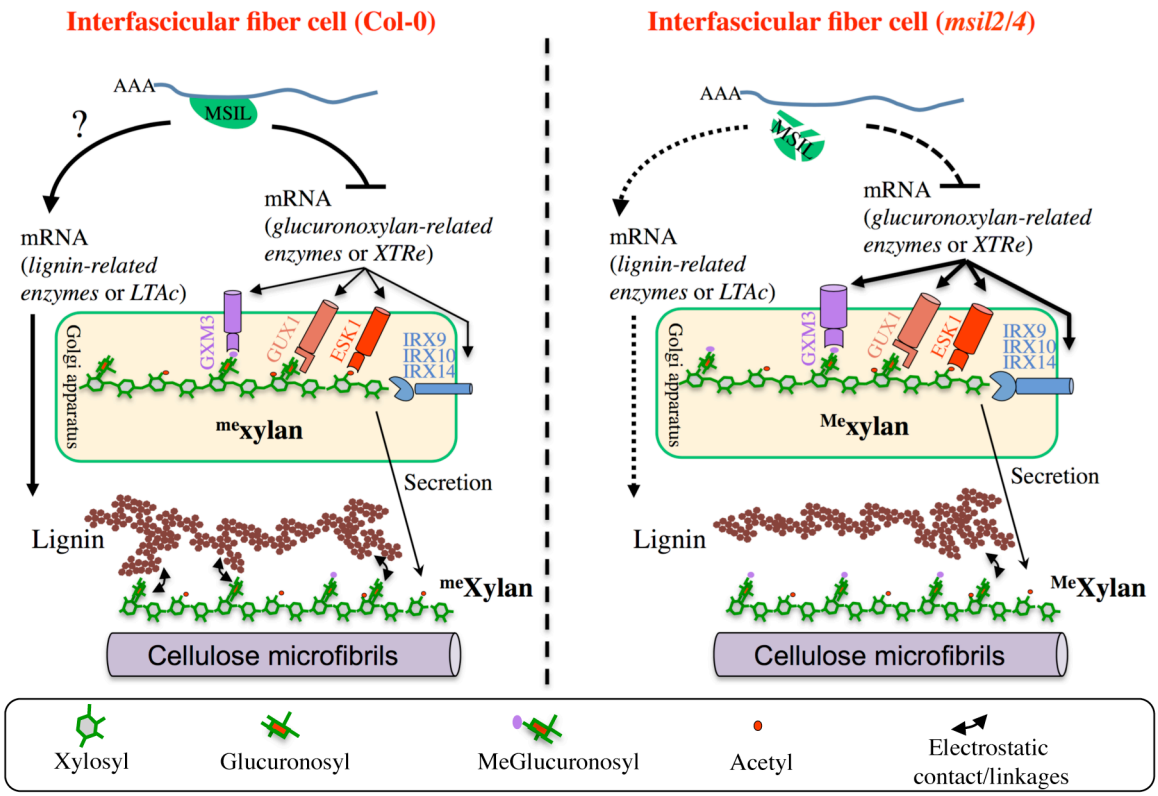


Fig. 7| Model of MSIL-dependent control of glucuronoxylan methylation in Arabidopsis and its consequence for SCW architecture. In the interfascicular fiber cells of Col-0, MSIL2/4 restrain the translation of the glucuronoxylan biosynthesis machinery, including the rate-limiting GXM3 enzyme. This activity would keep the level of glucuronoxylan methylation at an intermediate level, therefore providing a biochemical environment that favors the interactions between the glucuronoxylan and lignin polymers. In the interfascicular fiber cells of *msil2/4* mutant, the translation of the glucuronoxylan biosynthesis machinery, including GXM3, is increased, leading to the deposition of an over-methylated form of glucuronoxylan that has detrimental effects on glucuronoxylan-lignin interaction and the SCW formation. In a non exclusive manner, MSIL2/4 could also have a positive role in lignin synthesis, whose defect in *msil2/4* would lead to a decrease in lignin content. The model was inspired from Grantham et al[37].