LC-MS-based proteomics study of the *Arabidopsis thaliana* nuclear proteome during pattern triggered immunity

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I

– Biowissenschaften –

der Martin-Luther-Universität

Halle-Wittenberg

vorgelegt

von Herrn Mohamed Adel Abdelaziz Ayash

geb. am 19.08.1985 in Kairo

verteidigt am 28.03.2023

Reviewer 1: Prof. Dr. Ingo Heilmann

Reviewer 2: Prof. Dr. Klaus Humbeck

Reviewer 3: Prof. Dr. Till Ischebeck

Table of Contents

Index of abbreviations

BSA	Bovine serum albumin
Cpf_n	Cellular enrichment score
C	Cycloheximide treated
DDA	Data-Dependent Acquisition
FP	Fluorescent protein
dNTP	Deoxyribonucleotide triphosphate
F	Flg22 treated
FC	Flg22 and cycloheximide treated
Flg22	22 amino acid N-terminal epitope of bacterial flagellin
HRP	Horseradish peroxidase
LFQ	Label free quantification
LC-MS	Liquid chromatography - Mass spectrometry
Nlp20	20 amino acid peptide (nlp20), which is a distinctive part of the necrosis and ethylene-inducing peptide 1 (Nep1) like protein
Npf_n	Nuclear enrichment score
NLS	Nuclei localization signal
NE	Nuclear envelope
ORF	Open reading frames
PQI	Protein quantification index
PRRs	Pattern recognition receptors
PAMPS / MAMPS	Pathogen or microbe associated molecular patterns
PTI	Pattern triggered immunity
PSM	Peptide spectral match
RT	Room temperature
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TFs	Transcription factors
T	Transgenic plants (T) expressing UBQ10pro: TurboID-YFP-NLS
TBS	Tris-buffered saline
TBST	Tris-buffered saline +Tween
TEMED	Tetramethylethylenediamine
WT	Wild type
W	Water treated control

List of tables

List of figures

Appendix 2 (supplementary figures)

1. Introduction

1.1 Proteomics

Innovative analytical technologies have made it possible to conduct discovery-driven studies on equal footing with hypothesis-driven studies (Yu et al., 2010). Discovery driven investigations are concerned with broad questions about proteins, genes or transcripts in an organism, cell, tissue or an organelle. On the contrary, hypothesis driven investigations focus on a distinct pre-formed question about a single gene, transcript, protein or a smaller set of these (Yu et al., 2010). Development in novel techniques and instruments were the driving force to bring discovery-driven research to maturity. Development in cloning and gene sequencing facilitated the first steps towards studying the human genome, which was followed by evolution of transcript array analysis (Yager et al., 1991, Gilham, 1970).

Scientists started to focus on proteins shortly after these achievements in genomics and transcriptomics and after recognition of the open reading frames (ORF) as the draft of potential gene products (Ferguson and Smith, 2003, Tyers and Mann, 2003). Proteomics thus became a new field of science which was defined as the analysis of the entire proteome of an organism, tissue, cell or subcellular compartment under specific and several conditions (Ferguson and Smith, 2003). In contrast to genomics, proteomics had to face many difficult challenges for example, to resolve the extensive dynamic range of proteins, their post translational modifications, protein sample degradation, protein solubility and denaturation, alternative splicing products and other technical challenges (Tyers and Mann, 2003, Aebersold and Cravatt, 2002, Cho, 2007).

1.1.1 Proteomics development

Early steps in protein analysis were initiated by Pehr Edman in 1950. Edman introduced a method to sequence proteins by identification of residues at the N-terminal side of a protein or a polypeptide (Edman et al., 1950). An alternative approach was described by Sanger in 1959, enabling to deduce the insulin amino acid sequence through characterization of overlapping peptides (Sanger, 1959). The Edman method was later partly automated and was used extensively for a long time (Edman and Begg, 1967), but the method had disadvantages of

being laborious, time consuming and requiring large amount of protein sample. In the beginning of 1990s, mass spectrometry, a more sensitive and faster technique, replaced the Edman degradation method.

1.1.1.1 Mass spectrometry and ionization theory

The mass spectrometry instruments measure the mass to charge ratio (m/z) of the ionized analytes in the gas phase. In the classical definition, a mass spectrometer consists of 3 main parts: an ion source that ionizes the analyte molecule, a mass analyzer that measures and separates the ionized analytes according to their m/z ratio and a detector that records the ions count for each m/z value (Han et al., 2008). The two most commonly used soft ionization techniques are electrospray ionization (ESI) (Fenn et al., 1989) and the matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988) (Figure 1). In proteomics, there are four main types of mass analyzers. The ion trap, quadrupole, time of flight and Fourier-transform ion cyclotron. They all differ in their physical design and analytical performance (Aebersold and Mann, 2003). Each of them is either used alone or in combination with others to combine their advantages.
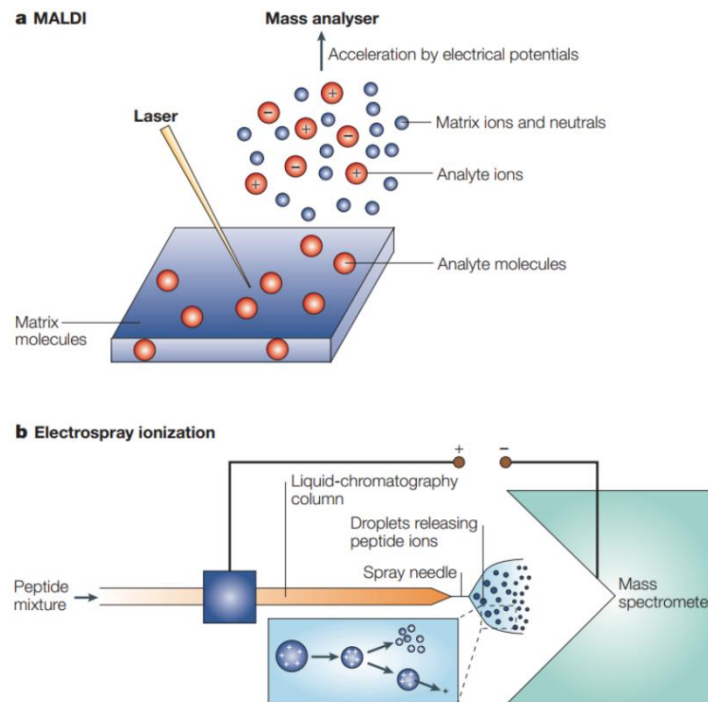
Figure 1. Ionization methods (Steen and Mann, 2004) (License number: 5371350041526, publisher: Springer Nature). (a) The matrix assisted laser desorption/ionization (MALDI). (b) The electrospray ionization (ESI). Details discussed in text.

1.1.1.2 Separation techniques coupled to mass spectrometry

The first two-dimensional electrophoresis gels (2-DE), in 1969 and 1970, enabled the separation of proteins or proteins complexes (Macko and Stegemann, 1969, Kaltschmidt and Wittmann, 1970). More importantly, higher analytical resolution was achieved by combining isoelectric focusing and SDS polyacrylamide gel electrophoresis (O'Farrell, 1975). The 2-DE coupled with MS (mass spectrometry) became an important tool in proteomics. Commonly, 2-DE is used to separate proteins followed by digestion of separated protein spots into peptides, which are then analyzed by MALDI–MS. In MALDI the analytes are mixed with crystalline matrix and then subjected to a laser beam which sublimes and ionizes the samples and finally the ions are detected by a mass analyzer (Figure 1) (Steen and Mann, 2004). This technique was used in the middle of the nineties to analyze proteins extracted from *Escherichia coli* and yeast (Henzel et al., 1993, Shevchenko et al., 1996). Despite the success of this technique in proteome research, it has several limitations. For example, 2-DE had problems with sensitivity, throughput, reproducibility and separation of complex protein mixtures (Erhardt et al., 2010). In

addition, MALDI lacked the on-line coupling to LC and is used typically for analysis of simple peptide mixtures (Steen and Mann, 2004). Liquid chromatography (LC) was introduced to overcome the limitations of 2-DE. Digested peptide samples are separated by nanoscale capillary high performance liquid chromatography (nHPLC) and are eluted from analytical columns according to their hydrophobicity. Then, the peptides flow through a needle tip and are vaporized and ionized (ESI) (Fenn et al., 1989). This process is called LC-MS which is the most commonly used approach in proteomics (Figure 1). The second step, after the ionization of the peptides, couples two stages of MS and is called tandem MS (MS/MS) (McLafferty, 1981). In tandem MS, the ionized peptides are separated in the first mass analyzer and then the separated or selected peptide ions (precursor ions) are fragmented in a collision cell by a technique called collision-induced dissociation (CID) (Shukla and Futrell, 2000). Other methods of fragmentation are also used for example, electron-capture dissociation (ECD) (Zubarev et al., 1998, Zubarev, 2006) and electron-transfer dissociation (ETD) (Syka et al., 2004, Pitteri et al., 2005). Fragment ions are then analyzed by the second mass analyzer creating the MS/MS spectra.

1.1.2 Proteomics approaches for protein identification and quantification

1.1.2.1 Bottom-up and top-down proteomics

Bottom-up and top-down proteomics are the main two strategies for protein identification by mass spectrometry. The main difference is that in bottom-up approaches the digested peptides are analyzed on the MS. By contrast, in top down approaches the intact proteins are directly analyzed by MS. The bottom-up strategy is the most commonly used approach for large scale analysis of extremely complex samples and it is divided into two main workflows (Han et al., 2008) (Figure 2). In the first workflow, the proteins are separated to produce a single or less complex protein mixture and then they are digested into peptides. Afterwards, the peptides are either directly analyzed by MS (Henzel et al., 1993) or undergo more peptide separation by LC coupled to tandem mass spectrometer (Loo et al., 2005). The second more popular workflow is named shotgun proteomics (McDonald and Yates 3rd, 2003, Cox and Mann, 2011). In contrast to the first workflow, the proteins are directly digested into peptides which can be separated by high performance liquid chromatography followed by tandem mass spectrometry. In the final step the generated MS/MS spectra are searched against the database of proteins through various programs like MASCOT (Perkins et al., 1999) or SEQUEST (Eng et al., 1994). Several

algorithms are run to match the identified spectrum with theoretically generated spectral peak lists to identify the peptides, then the identified peptides are used to infer protein identifications. In top-down proteomics, the proteins are separated and then the intact proteins are ionized and subsequently analyzed by tandem mass spectrometry (McLafferty et al., 2007) (Figure 2). On one hand, this new approach enabled complete reach and characterization of protein sequence including post translational modifications (PTM) and removed the tedious step of protein digestion when compared to the bottom-up approach. On the other hand, limitations in analytical throughput, proteome sensitivity and data analysis hampered the use of this strategy (Kellie et al., 2010, Catherman et al., 2014).
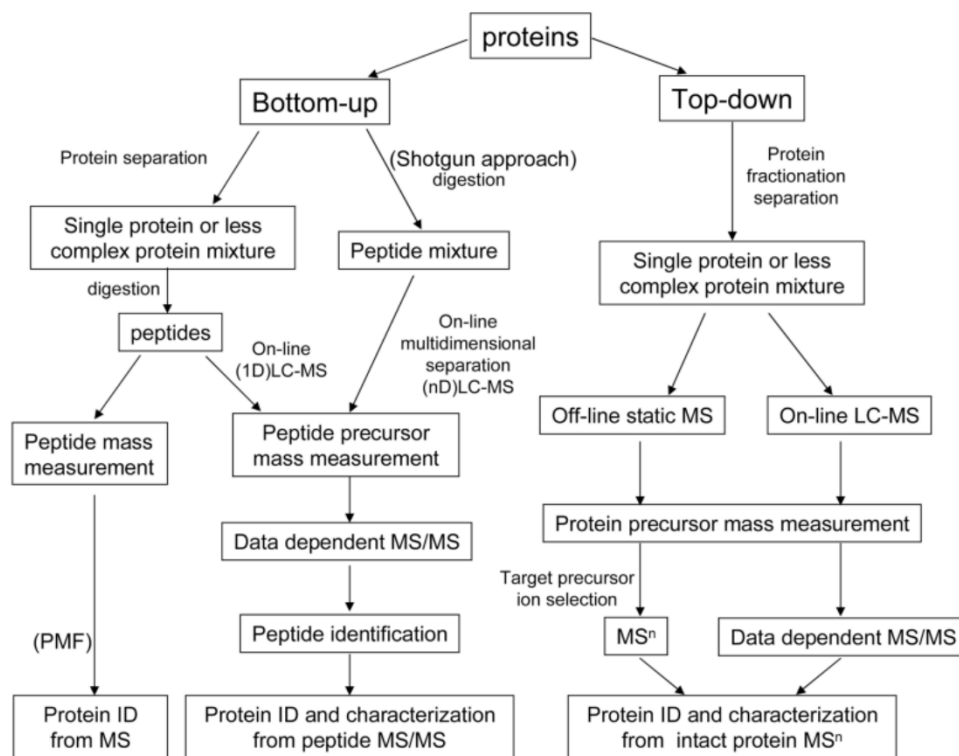
Figure 2. Bottom-up and top-down proteomics (Han et al., 2008) (License number: 5371331420276, publisher: Elsevier). Scheme showing the different strategies of MS-based proteomics. Details discussed in text.

1.1.2.2 Discovery and targeted proteomics

Discovery proteomics main objective is to identify and quantify the entirety of proteins in a biological sample. It is commonly used when the focus is the sample breadth and large scale

protein coverage (Hu et al., 2016). The most common analysis mode used in discovery proteomics is data-dependent acquisition (DDA). In the DDA scan strategy, the most abundant peptide precursor ions on MS1 are isolated and then fragmented and analyzed with MS2 (Figure 3) (Mann et al., 2001, Bateman et al., 2014). Targeted proteomics is used for precise and reproducible quantification of few selected proteins in a biological sample (Hu et al., 2016, Chen and Liu, 2019, Marx, 2013). Targeted proteomics allows only limited number of data points over a peak and therefore, limits the number of peptides analyzed (Gallien and Domon, 2015). This can be expanded to around a few hundred in a single run, if the retention time scheduled windows approach is used (Gallien et al., 2012, Majovsky et al., 2014). In contrast to DDA, targeted approaches select predetermined precursor peptide ions in the MS1 scan. Afterwards, the selected precursor ions are fragmented and a subset of fragment ions are analyzed on MS2 in a multiple reaction monitoring analysis mode (MRM) (Lange et al., 2008) or all peptide fragment ions are measured by MS2 in a parallel reaction monitoring analysis mode (PRM) (Figure 3) (Peterson et al., 2012). Advances in the mass spectrometers and bioinformatics programs allowed the development of a new analysis method called data-independent acquisition (DIA). The goal of this approach is to combine the identification broadness of DDA with the precise and reproducible quantification of MRM/PRM. In DIA mode, all precursor ions, from a specific mass range on MS1 scan, are fragmented and analyzed in MS2 (Figure 3) (Gillet et al., 2012, Bilbao et al., 2015, Venable et al., 2004). Despite that DIA is considered as the next generation proteomics technique, there are some limitations that need to be tackled, for instance: deconvoluting the complexity of MS2 spectra and proper filtering for false positives and false negatives. Also, DIA requires many experiments to collect spectral libraries for peptide identification which is a major problem in the case of subcellular proteomics where the subcellular proteome is undervalued in a reference spectral library (Krasny and Huang, 2021, Hu et al., 2016). The future prospective of DIA is positive as the improvements in hardware, bioinformatics software and data learning is ongoing and some of the limitations have already been overcome (Hu et al., 2016).
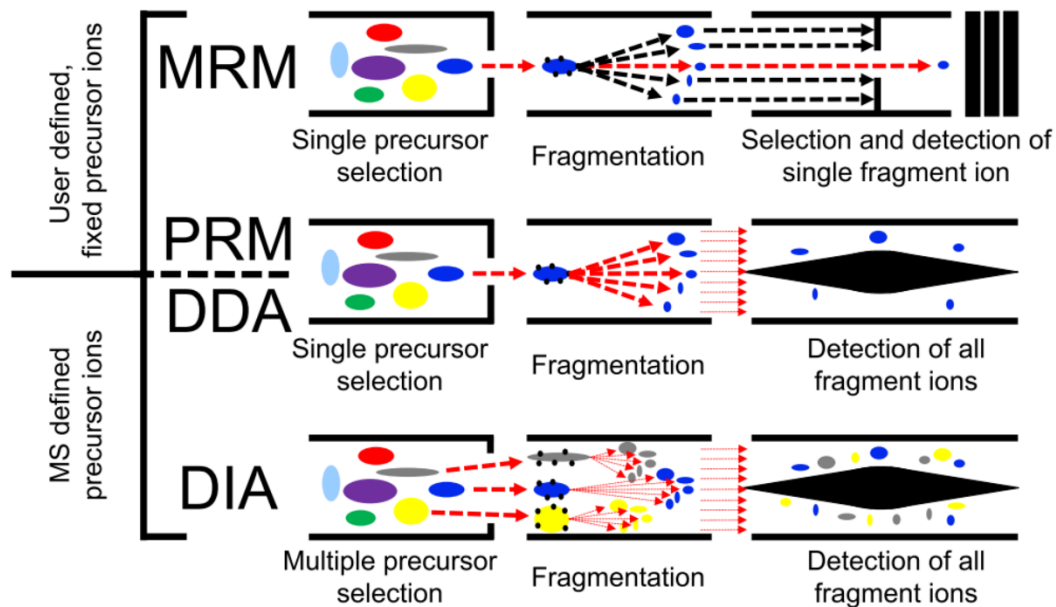
Figure 3. Proteomics analysis modes (Hu et al., 2016). licensed under CC BY 4.0. In DDA, top (n) precursor ions are selected, fragmented and analyzed in MS2. In PRM, selected precursor ions are isolated, fragmented and then analyzed in MS2. In MRM, selected precursor ions are isolated, fragmented and selected fragment ions are analyzed in MS2. In DIA, all precursor ions, from a specific mass range on MS1 scan, are fragmented and analyzed in MS2.

1.1.3 Quantitative proteomics

Quantitative proteomics can be categorized into two main approaches: label-based and label-free methods (Bantscheff et al., 2012). In label-free approaches, the peptides and proteins are quantified without introducing any kind of labels. It is divided into two fundamental approaches (Figure 4): spectral counting and ion intensity based quantification (XIC) (Neilson et al., 2011, Ong and Mann, 2005). Spectral counting is based on the number of peptide spectral matches (PSM), where the most abundant peptides will be selected and fragmented then analyzed with MS2 and therefore will produce higher numbers of MS/MS spectra which is proportional to the protein abundance in the sample (Liu et al., 2004). In contrast to spectral counting, ion intensity based quantification is based on measuring the chromatographic peak area of the eluted peptide. The measured area, extracted ion current XIC, is linearly proportional to the measured peptide (Chelius and Bondarenko, 2002, Bondarenko et al., 2002). Old et al., compared the label free methods and their findings state that, spectral counting is validated to be more sensitive with proteins that are changing in abundance, while the intensity based quantification

gave more precise assessments of protein ratios (Old et al., 2005). Moreover, spectral count variability of low abundance proteins hindered the sensitivity of the quantification of these proteins (Lee et al., 2019a).

The label-based approach is divided into two methods (Figure 4): metabolic labeling and chemical labeling. Metabolic labeling is based on the concept of using stable heavy isotopes to label peptides and is best exemplified by stable isotope labeling with amino acids in cell culture (SILAC) (Ong et al., 2002). In this method, isotopic labeled amino acids (heavy) are added to the cell culture media at an early stage of cell growth and differentiation, whereas, unlabeled cell culture media is supplemented with light amino acids. Finally, the tissues from both samples are mixed, the protein extract is digested, analyzed by LC-MS and the labeled and unlabeled peptides are relatively quantified. Later pulsed SILAC was introduced to measure protein turnover rates (Milner et al., 2006, Lam et al., 2007). Chemical labeling quantification was first introduced with the ICAT method (isotopic coded affinity tags) (Gygi et al., 1999). In this method the ICAT reagents with heavy and light linkers are added to cysteine side chain residues of proteins by forming a thiol-ester bond in different samples, subsequently allowing relative quantification of labeled to unlabeled proteins. Later, TMT (tandem mass tags) and iTRAQ (isobaric tags for absolute and relative quantification) isobaric methods were introduced (Thompson et al., 2003, Ross et al., 2004, Wiese et al., 2007). In isobaric methods, isobaric labels are added to the N- terminal side of proteins/peptides and the lysine on side chains. The same peptides labelled with different labels have identical mass but after fragmentation, they are differentiated by detection and quantification of different reporter ions.

Label free approaches were introduced to overcome many of the drawbacks of the label-based methods, some of which were expensive cost, requirements of specific software for data analysis, limited number of samples per experiment and no applicability to all types of samples (Neilson et al., 2011). Label free approach gained popularity especially after the introduction of hybrid linear ion trap/orbitrap mass spectrometers which are characterized by high resolution and high mass accuracy analysis (Makarov et al., 2006). However, label free methods have a major disadvantage that the samples are analyzed separately which could lead to technical error between samples. This problem was avoided by running all samples in the same sequence and on the same instrument followed by data normalization (Bantscheff et al., 2012). In addition, many softwares and statistical methods were used to overcome these variabilities. Finally, development in data analysis software and search engines, such as MaxQuant (Cox et al., 2014), Proteome Discoverer (Thermo Fisher), MASCOT (Perkins et al., 1999) and

SEQUEST (Eng et al., 1994), promoted the quantitative analysis of several thousands of proteins by label free methods.
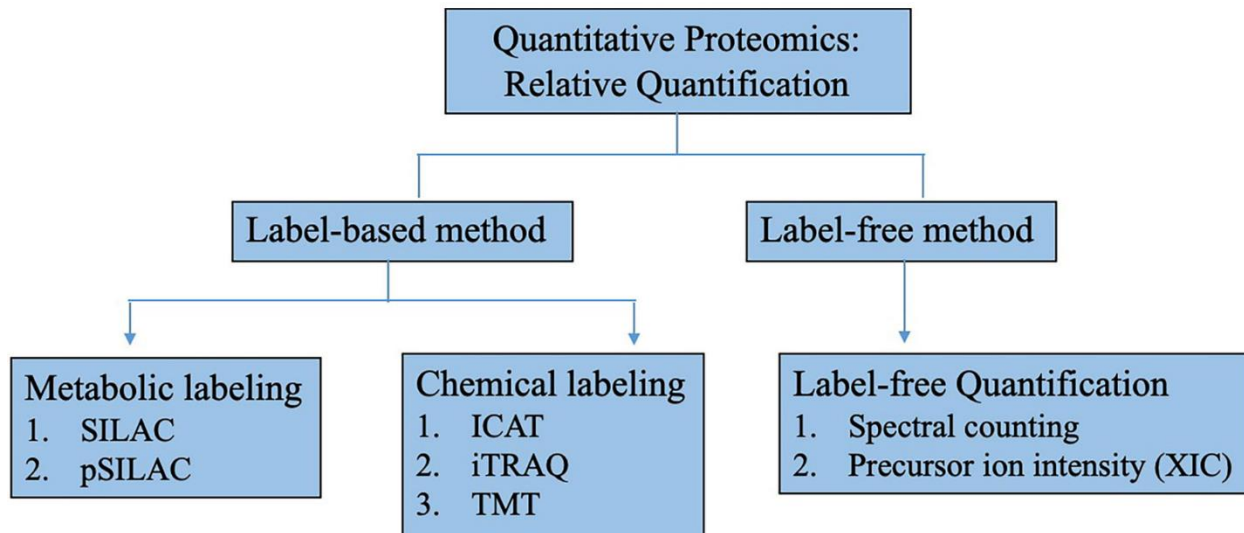
Figure 4. Methods for quantitative proteomics (Vinaiphat et al., 2021). licensed under CC BY 4.0. Quantitative proteomics was divided into label-based and label free methods. Label free methods are based on two approaches: SC (spectral count) and XIC (precursor ion intensity). Labeling is done either chemically: ICAT (isotopic coded affinity tags), iTRAQ (isobaric tags for absolute and relative quantification) and TMT (tandem mass tags) or metabolically: SILAC (stable isotope labeling with amino acids in cell culture) and pSILAC (pulsed SILAC).

1.2 Subcellular protein localization

The two main tools to determine protein localization in the plant cell are: fluorescent protein (FP) tagging and subcellular proteomics. FP tagging is commonly used when the question is to determine the localization of specific proteins of interest. However, subcellular proteomics is mainly used when the whole proteome of certain subcellular organelle is investigated (Tanz et al., 2013).

1.2.1 Subcellular proteomics

Subcellular compartmentalization is a key feature of eukaryotes. Organelles are specialized structures found in the plant cell and are interacting with each other and the cytosol (Agrawal et al., 2011). Different sets of proteins are found in the subcellular organelles and their localization defines their function. Subcellular proteomics also known as organellar proteomics, is the

investigation of the proteome at the subcellular compartments level (Christopher et al., 2021). Mainly subcellular proteomics is performed by the isolation of the organelles in question followed by total organelle protein extraction and MS analysis. An alternative newer approach is proximity dependent protein labeling coupled with MS, where proteins are labeled in their compartments without isolation followed by analysis of labeled proteins.

1.2.1.1 Organelle isolation

Organelle isolation procedures are divided into two main steps: (1) cell wall and membrane lysis and (2) fractionation of the crude extract (Agrawal et al., 2011). Cell wall and membrane disruption are performed on the plant tissue by homogenization, which can be carried out by razor blade or by grinding under liquid nitrogen. The homogenizing medium usually consists of Tris or HEPES buffers to control the pH, reducing agents like DTT (dithiothreitol), protease inhibitors and non-ionic detergents like Triton x-100 to help in organelle release (Loureiro et al., 2006, Loureiro et al., 2007). Another method is to prepare the plant protoplast which is gentler than homogenization but requires extra work. In fractionation, the first step is often a filtration step by nylon mesh to remove large contaminants. Secondly, differential centrifugations are done to isolate the target compartment and separate it from other organelles. Finally, the enriched compartment can be purified by percoll density gradient centrifugation (Yin and Komatsu, 2016, Folta and Kaufman, 2006, Kumar et al., 2014, Bae et al., 2003). Organelle enrichment is always accompanied by inevitable contaminations and the success of the method depends on the purity of isolated organelles. LOPIT (localization of organelle proteins by isotope tagging) is a method that was established to partially isolate organelles by density gradients, then the protein partitioning in the different gradients is measured by ICAT labeling and mass spectrometry (Dunkley et al., 2004). This method is independent on the purity of the organelles. Nevertheless, the method is less sensitive to low abundant compartment proteins (Agrawal et al., 2011). Also, the method could lead to wrong interpretations, for example identifications of cytosolic proteins in mitochondria would claim impurity of mitochondrial preparation as done by (Giegé et al., 2003).

1.2.1.2 Proximity-dependent labeling

Contamination of isolated organelle, which could lead to false positive identification, is the main drawback of the organelle isolation method. Therefore, proximity dependent labeling was developed to evade this problem. Generally, proximity labeling is based on certain enzymes

which are able to add biotin to nearby interacting proteins or to proteins in a certain neighborhood (Chen and Perrimon, 2017, Li et al., 2017) (Figure 5). The enzymes are usually fused to signal peptide to target the enzyme to certain compartment in order to biotinylate the proteins inside it. Also, the enzymes could be fused to protein of interest to biotinylate the interacting proteins in its proximity. After the proximity labeling is executed, the total proteins are extracted and the biotinylated proteins are isolated using streptavidin beads. Next, the biotinylated proteins are digested and generated peptides are analyzed by mass spectrometry (Bosch et al., 2021) (Figure 5). One of the approaches that utilizes proximity labeling is called BioID. The prototype of BioID was the biotin ligase BirA which biotinylates proteins containing BAP (biotin acceptor peptide) (de Boer et al., 2003). Subsequently, the active site of BirA was mutated to generate BioID, which is able to promiscuously label proteins without BAP (Roux et al., 2012). BioID2, a newer version of BioID, was developed to improve the activity and decrease the amount biotin required (Kim et al., 2016). The BioID system has a few disadvantages, for example a long labeling time from 15 to 24 h and reduced activity below 37 degrees (Chen and Perrimon, 2017). Another essential point that until recently the applications of BioID in plants was limited to few studies using transient and elevated expression (Khan et al., 2018, Conlan et al., 2018, Lin et al., 2017). In 2018, TurboID was developed and overcame the BioID drawbacks for instance, it has higher activity, reducing labeling time to about 10 min for western blot detectable signals and to a few h for MS identification and it works at room temperature (RT) (22 degrees) (Branon et al., 2018). Additionally, TurboID was validated to effectively work under various conditions, expression levels and different developmental stages and tissues in plants (Zhang et al., 2019, Mair et al., 2019).
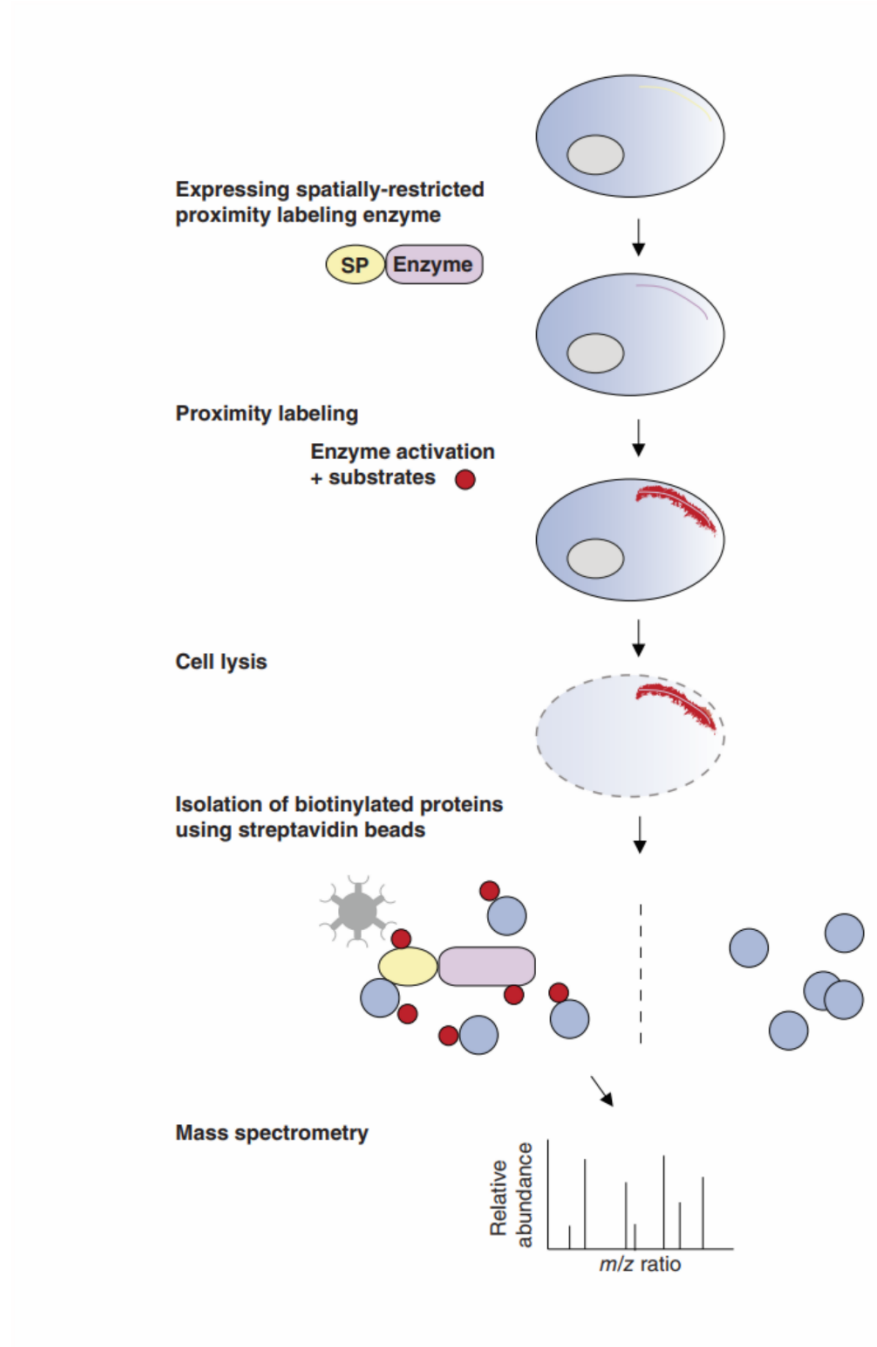
Figure 5. Proximity dependent labeling (Chen and Perrimon, 2017) (License number: 5371340642503, publisher: John Wiley and sons). Enzymes are fused to signal peptides (SP) to direct them to certain subcellular compartments. After expression they biotinylate the proteins in this compartment. Finally biotinylated proteins are enriched on streptavidin beads and digested for MS analysis. Enzymes could be fused to proteins of interest (SP) to label the interacting proteins in its proximity.

1.2.1.3 Plant nuclear proteomics

The plant nucleus is a vital and prominent organelle in the plant cell. It contains various types of DNA and RNA, nuclear bodies, sub-compartments and proteins. The plant nucleus is surrounded by nuclear the envelope subsisting of two nuclear membranes: the outer nuclear membrane and the inner nuclear membrane. Both membranes merge at the nuclear pore and contain the nuclear pore complexes, which allow protein import into the nucleus (Meier et al., 2017). While the protein constituents of the inner nuclear envelope are not well characterized in plants, it is speculated that structures similar to the inner lamina of mammalian nuclear envelopes also exist in plants.The plant nuclear lamina-equivalent likely contains lamin-like proteins, such as the nuclear matrix constituent protein family (NMCP1) (Ciska and Moreno Díaz de la Espina, 2013). LINC (little nuclei) proteins likely form complexes acting as connecting points between the lamina-equivalent and the cytoskeleton (Zhou et al., 2012), as has been proposed for human nuclei (Sosa et al., 2012). Chromatin is attached to inner nuclear membrane and is scaffolded on histone proteins which are facultative in the nucleus of all eukaryotes and have a mass almost equal to the mass of DNA (Wiśniewski et al., 2014). During cell division, the nuclear envelope disassembles and the mitotic spindles gets access to the chromosomes. This is always associated with changes in chromatin structure and chromosome arrangement (Kutay and Hetzer, 2008).

Several processes like DNA replication, transcription regulation and epigenetic regulation occur in the nucleus. These functions require many molecules like proteins (Petrovská et al., 2015). Proteins are considered to be the most abundant constituent of the nucleus (Sutherland et al., 2001) . Therefore, the investigation of the nuclear proteins is crucial to understand the function of the processes occurring in the nucleus. Generally, the plant nuclear proteome is understudied when compared to the mammalian and human nuclear proteome (Thul et al., 2017, Go et al., 2021, Petrovská et al., 2015, Yin and Komatsu, 2016, Jez et al., 2021). There are two general approaches to define the plant nuclear proteome. The first one is based on deducing unknown plant nuclear proteins from known proteins from other species (Petrovská et al., 2015). The progress in this ortholog based approach is limited because of several factors such as: the significant differences in the nuclei between plants and mammals (Meier, 2009), around 18% of identified plant nuclear proteins are dissimilar to proteins characterized in other organisms (Narula et al., 2013) and the large number of nuclear proteins hindered progress (Petrovská et al., 2015). As a result, mass spectrometry based nuclear proteomics is the most commonly used method to study the nuclear proteome by way of plant nucleus isolation, protein extraction

and then analysis by gel-free LC/MS or gel based 2-DE/MS (Erhardt et al., 2010) (Figure 6). Furthermore, proximity labeling was employed recently for organellar protein localization as discussed above. The *Arabidopsis thaliana* nuclear proteome was first investigated in 2003 using two dimensional electrophoresis (2-DE) (Bae et al., 2003) . Afterwards, another six *A. thaliana* studies reported sub-nuclear proteomes for instance: nucleolus, nuclear matrix, nuclear envelope and chromatin (Calikowski et al., 2003, Pendle et al., 2005, Sakamoto and Takagi, 2013, Bigeard et al., 2014, Chaki et al., 2015, Tang et al., 2020). More recently, four studies have defined the core *A. thaliana* nuclear proteome by LC-MS comprehensively (Palm et al., 2016, Goto et al., 2019, Mair et al., 2019, Ayash et al., 2021). The latter study is equivalent to the first part of the results in this thesis.
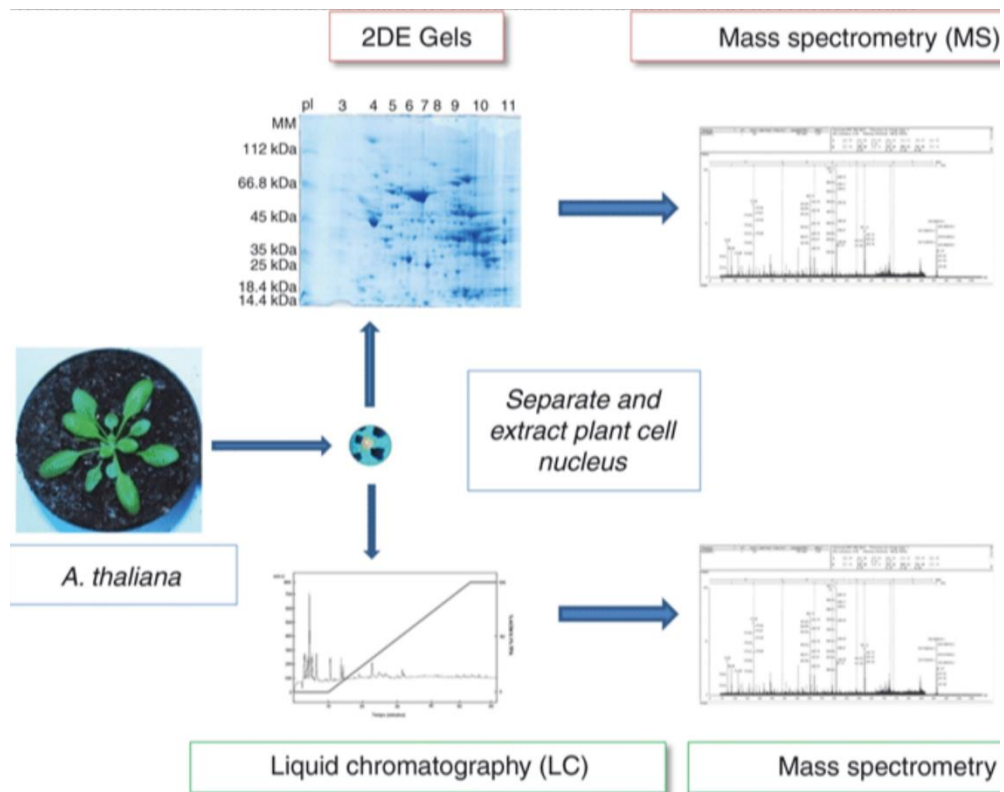
Figure 6. Scheme representing the gel-free vs. gel-based mass spectrometry approaches to analyze the plant nuclear proteins (Erhardt et al., 2010) (License number: 5371340783221, publisher: John Wiley and sons) modified.

1.3 Nuclear Import, dual targeted proteins and protein trafficking

Nuclear pore complex (NPC) consists of more than thirty nucleoporins which makes it one of the largest multiprotein complexes in *Arabidopsis thaliana* (Tamura et al., 2010, Tamura and Hara-Nishimura, 2013). NPCs act as barrier for the transport of molecules in and out of the nucleus. The transport of molecules across the nuclear pore is controlled by two different processes: passive diffusion and active transport (Di Ventura and Kuhlman, 2016). The passage of large macromolecules is regulated by the active transport process and requires the assistance of nuclear transport receptors such as importins and exportins (Bednenko et al., 2003, Mosammaparast and Pemberton, 2004, Stewart, 2007). The transport receptor binds the cargo and forms the transport receptor complex which is then translocated by the NPC through the nuclear envelope. Importin α comprises two domains, the first one is the C-terminal domain called ARM which binds the nuclear localization signal (NLS) on the macromolecule and the second N-terminal domain binding the importin β (Goldfarb et al., 2004, Marfori et al., 2011). After the importin α has bound both the macromolecule and importin β, the importin β interacts with Ran GTPase and allow the nuclear import of macromolecules (Lott and Cingolani, 2011, Merkle, 2011). In contrast to large molecules, it is known that proteins smaller than 40-60 kDa can passively diffuse through the barrier (Christie et al., 2016, Mohr et al., 2009, Keminer and Peters, 1999, Ma et al., 2012, Weis, 2003, Ribbeck and Görlich, 2001, Görlich, 1998, Timney et al., 2016), but also proteins larger than 60 kDa were shown to diffuse in a passive manner through the NPC (Wang and Brattain, 2007). Another study introduced a different mechanism for the passive diffusion through NPCs, known as soft barrier which replaces the traditional rigid size barrier (Timney et al., 2016). In this new model a soft barrier for the passive diffusion becomes increasingly stronger with the increasing molecular mass of diffusing molecules. In addition to the classical NLS and importin mediated pathways, alternative mechanisms for nuclear transport were also investigated (Guinez et al., 2005, Imamoto and Kose, 2012, Sobočanec et al., 2016).

Dual targeted proteins (dual localized) are a number of proteins which can be alternatively located in different organelles. Dual targeting has been reported for proteins alternatively residing in mitochondria or plastids (Sharma et al., 2018), nucleus or plastids (Schwacke et al., 2007) and nucleus or mitochondria (Carrie et al., 2009a). Dual targeted proteins can have alternative functions in different organelles or one main function in one targeted organelle, and it was suggested that dual targeting could be an evolutionary mechanism to establich biochemical pathways in parallel in different compartments (Martin, 2010). Generally, dual localization can

be classified into two main categories: dual targeting of newly synthesized proteins and relocalization of mature proteins from one organelle to another (Krause and Krupinska, 2009, Krupinska et al., 2020). In the first category of *de novo* synthesized proteins, different strategies could be utilized for instance: formation of multiple proteins from (i) different transcription start sites, (ii) translation start sites and (iii) alternative splicing (Sunderland et al., 2006, Ohta et al., 1995) and another strategy is the post-translational modification of individual protein (Galichet et al., 2008). Dual targeted proteins could have different targeting sequences or the targeting sequence may be ambiguous (Peeters and Small, 2001). Relocation and trafficking of proteins and other molecules from one compartment to another plays a role in the communication and dynamics of plant cell compartments. This communication is necessary to co-ordinate their activities and functions. Two main communication and signaling processes occur in the cell: the first one is anterograde signaling where the nucleus regulates the functions of other organelles like chloroplast and mitochondria, whereas the second is called retrograde signaling which is the back flow of instructions from the organelles to the nucleus reporting their functional state (Bräutigam et al., 2007). Retrograde signaling and subcellular protein trafficking are vital players in the cellular response to biotic and abiotic stresses (Kmiecik et al., 2016, Crawford et al., 2018, Bobik and Burch-Smith, 2015).

1.4 Plant immunity

The first line of defense for plants against potentially harmful pathogens and microbes is the physical barrier for example: the cell wall (Malinovsky et al., 2014), lignin (Lee et al., 2019b, Sattler and Funnell-Harris, 2013), cuticle (Yeats and Rose, 2013) and formation of papillae at sites of infection (Underwood, 2012). In addition, chemical barriers like toxins and antimicrobial secondary metabolites are also used by plants (War et al., 2012, Bednarek et al., 2009). When the plant physical barriers fail against the pathogens, the plant initiates its innate immune response. The plant immunity is classified into: pattern triggered immunity (PTI) and effector triggered immunity (ETI) and is best described by the four phased zig-zag model (Figure 7) (Jones and Dangl, 2006). PTI is initiated by the recognition of the PAMPS / MAMPS (pathogen or microbe associated molecular patterns) and DAMPS (damage associated molecular pattern) by receptors found in the plasma membrane called pattern recognition receptors (PRRs) (Nishad et al., 2020, Saijo et al., 2018, Zhou and Zhang, 2020) such as the leucine-rich repeat receptor like kinases (LRR-RLKs) and receptor like proteins (LRR-RPs) which sense peptide patterns and lysine motif receptor or lectin S-domain receptor kinases sensing chitin and

bacterial peptidoglycan (Boutrot and Zipfel, 2017, Ranf, 2017, Gust et al., 2012, Miya et al., 2007). One of the best studied PAMPs is flg22, which represents an N-terminal 22 amino acid epitope of bacterial flagellin (Meindl et al., 2000). Flg22 is recognized at the surface of plant cells by the LRR-RLK flagellin sensing 2 (FLS2) (Gómez-Gómez and Boller, 2000). More recently, the 20 amino acid peptide nlp20, which is a distinctive part of the necrosis and ethylene-inducing peptide 1 (Nep1) like protein (NLP), was shown to trigger PTI after its recognition by the LRR-RP, RLP23 (Albert et al., 2015, Böhm et al., 2014). Early responses following PTI activation are associated with the production of ROS and a calcium burst. Calcium controls calcium dependent protein kinases (CPKs) which play a role in transcriptional regulation in immunity (Boudsocq et al., 2010). Another PAMP activated signal transduction cascade involves mitogen activated protein kinases (MAPKs) which are a main signaling module in immunity, culminating in the phosphorylation and activation of various substrates which regulate transcriptional reprograming (Bigeard et al., 2015). ROS production in PTI is regulated by phosphorylation of NADPH oxidase at RBOHD by receptor like kinase (BIK1), BIK1 phosphorylation is stimulated by FLS2 and BAK1 phosphorylation which happens directly after perception of flg22 by FLS2 (Macho and Zipfel, 2014, Lu et al., 2010, Kadota et al., 2014, Zhang et al., 2010). PTI is also regulated by various phytohormones such as jasmonic acid, salicylic acid and ethylene (Pieterse et al., 2012). In order to overcome PTI, the microbes delivers effector proteins into the plant cell resulting in effector triggered susceptibility (Jones and Dangl, 2006). Afterwards, the pathogen effectors are recognized by NLR (nucleotide binding-leucine rich repeat) intracellular receptors and induce effector triggered immunity (ETI) (Dangl and Jones, 2001). NLR can also recognize the host proteins modified by the pathogen effectors (Dodds et al., 2006, Ade et al., 2007, Wang et al., 2019). ETI always induces local resistance and localized cell death named the hypersensitive response (HR) (Jones and Dangl, 2006).
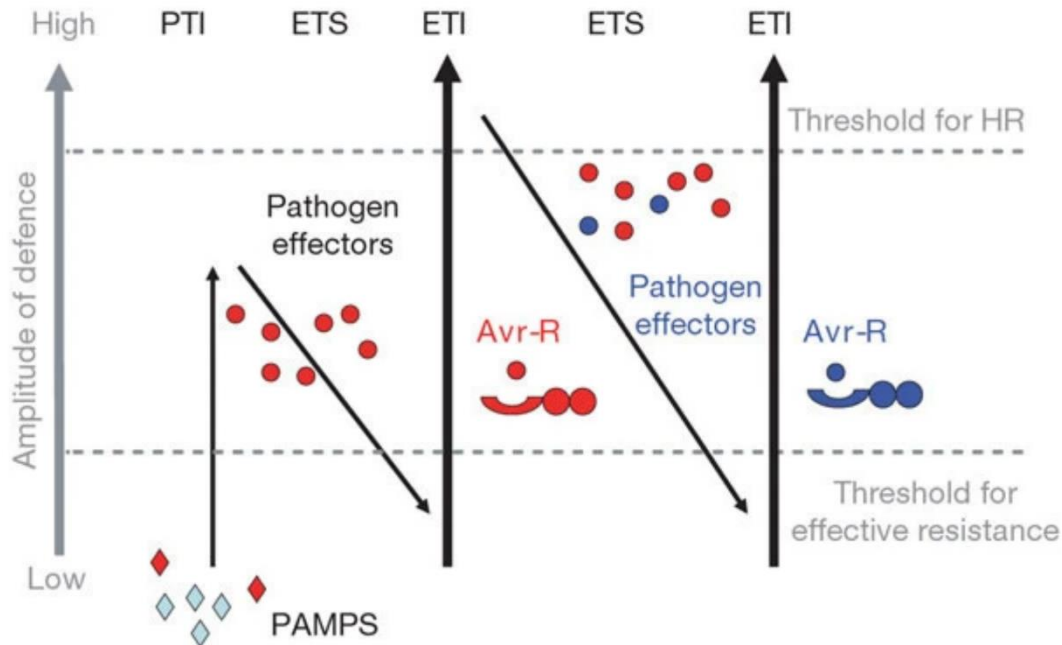
Figure 7. Zig-zag model of plant immunity (Jones and Dangl, 2006) (License number: 5371350180461, publisher: Springer Nature)**.** This model describes four phases of immunity. In the first phase, PAMPS and MAMPS are detected by the PRR and triggers PTI. In the second phase, microbes overcome PTI by the release of virulence effector proteins to the plant cell resulting in effector triggered susceptibility (ETS). In the third phase, the effector proteins are recognized by intracellular receptors activating ETI. In the last phase, pathogens release new effectors to suppress ETI and then plant recognize these effectors and trigger ETI again.

Several studies have shown that the nuclear localization of microbe's virulence effectors, R proteins, transcription factors, transcription co-activators, general regulators and others, are crucial for plant defense against infection (Rivas and Deslandes, 2013, Deslandes and Rivas, 2011). Moreover, many nuclear processes are involved in immunity such as nuclear transport, post translational modifications, histone and DNA modifications, transcription regulation, alternative splicing and RNA interference (Motion et al., 2015). Trafficking across the nuclear membrane, for mRNA (Gaouar and Germain, 2013) and effector proteins and their host resistance proteins (Bhattacharjee et al., 2013), has shown to be essential in plant immunity. Transcription regulation is a main player in immunity as it has been shown that around 20 - 25 % of genes undergo transcriptional changes and are regulated by high numbers of transcription factors upon infection (Raffaele and Rivas, 2013, Padmanabhan and Dinesh-Kumar, 2010). Also, the activation of many transcription factors often needs post translational modifications like phosphorylation (Park et al., 2012). Besides, it has been shown before in a transcriptomic study

that a significant number of *Arabidopsis thaliana* genes could undergo alternative splicing after *Pseudomonas* infection (Howard et al., 2013). The nuclear plant proteome in immunity is poorly investigated with only a few studies have been performed (to my knowledge) in tomato, potato, rice, soybean and apple leaves (Howden et al., 2017, Rajamaki et al., 2020, Narula et al., 2019, Cooper et al., 2011, Sikorskaite-Gudziuniene et al., 2017). Only one short communication paper investigated the *Arabidopsis thaliana* nuclear proteome in immunity after chitosan treatment (Fakih et al., 2016). Recently in 2021, our publication was the first to investigate the *Arabidopsis thaliana* nuclear proteome in PTI after flg22 and nlp20 treatments (Ayash et al., 2021).

1.5 Aims and objectives of the study

In spite of the vital role of the nucleus, the plant nuclear proteome remained understudied compared to the mammalian and human nuclear proteomes. Moreover, only few studies have investigated the core nuclear proteome of *Arabidopsis thaliana* comprehensively by LC-MS. In the context of plant immunity, very little is known about the *Arabidopsis* nuclear proteome including proteome rearrangement in pattern triggered immunity. The communication and interaction of plant cell organelles and how they coordinate their functions by trafficking of molecules and proteins, is a current focus of many research groups including our research training group.

My study aims to use LC-MS-based proteomic approaches to: (1) identify the *Arabidopsis thaliana* nuclear proteome, (2) investigate the nuclear proteome under biotic stress stimuli, (3) investigate the nuclear import including dual targeted proteins and trafficking from other compartments. In order to achieve the aim two main approaches were used. The first one was based on a nuclei isolation technique from *Arabidopsis* cell culture, followed by devising an enrichment score to assess the purity of the preparation and curate the nuclear proteome. The second one utilized proximity dependent labeling to circumvent the inevitable contaminations. Both of them were used to investigate the nuclear proteome in the frame of PTI. The objective is to present a high quality MS-draft catalogue of the *Arabidopsis thaliana* nuclear proteome that can complement the present knowledge and enrich it with potential putative candidates in the context of immunity and trafficking.

2. Materials & Methods

2.1 Materials

Chemicals, reagents & kits

The chemicals, reagents and kits used were purchased from Carl Roth, Thermo Fischer, Merck, Promega, Eppendorf, SERVA and Sigma-Aldrich unless another manufacturer is stated.

Table 1. Chemicals, reagents and kits

Name	Manufacturer
Antibodies	
Polyclonal anti-histone H3 antibody	Agrisera (Vännäs, Sweden)
Streptavidin-POD Conjugate (HRP)	Roche (Mannheim, Germany)
Anti-GFP polyclonal antibody (Rabbit)	Thermo Fisher (USA)
Anti-rabbit, HRP-linked Antibody	Bio-rad (Puchheim, Germany)
Kits	
Super signal west femto maximum sensitivity substrate	Thermo Fisher (Rockford, IL, USA)
C18 stage tips (in-house), Empore C18 extraction discs	3M (Neuss, Germany)
PD-10 desalting column, sephadex G-25 M (gel filtration)	GE Healthcare, life sciences (Cytiva) (Buckinghamshire, UK)
Protein Lo-Bind Tube	Eppendorf (Hamburg, Germany)
2-D Quant Kit	GE Healthcare, life sciences (Cytiva) (Buckinghamshire, UK)
Dynabeads Myone Streptavidin C1	Invitrogen (Vilnius, Norway)
Grinding micromilling beads	Mühlmeier (Baernau, Germany)
ECL prime Amersham western blotting detection reagent	Cytiva (Buckinghamshire, UK)
Pierce ECL plus western blotting substrate	Thermo Fisher (Rockford, IL, USA)
SV total RNA isolation system	Promega (Maidson, USA)
DNase I, RNase-free	Thermo Fisher (Vilnius, Lithuania)
Revert aid first strand cDNA synthesis kit	Thermo Fisher (Vilnius, Lithuania)
5x QPCR Mix EvaGreen®(Rox)	Bio&SELL GmbH (Nürnberg, Germany)
Amicon ultra 30 K filter	Millipore (Carrigtwohill, Ireland)
Other Reagents	
Acetonitrile ultra LC-MS	Carl Roth (Karlsruhe, Germany)
Ammonium persulfate	Sigma-Aldrich
Ammonium Acetate	Carl Roth GmbH (Karlsruhe, Germany)
Ammonium Bicarbonate	Carl Roth GmbH (Karlsruhe, Germany)
Acetone	Sigma-Aldrich (Steinheim, Germany)
Biotin	Sigma-Aldrich (Steinheim, Germany)

BSA (biotin free)	Carl Roth (Karlsruhe, Germany)
Blotting paper	Sartorius (Goettingen, Germany)
Blotting grade milk	Carl Roth (Karlsruhe, Germany)
Calcium chloride	Merck (Darmstadt, Germany)
CL-xposure film (blue x-ray film)	Thermo Fisher (Rockford, IL, USA)
Cycloheximide	Sigma-Aldrich (Steinheim, Germany)
Cellulase	SERVA (Heidelberg, Germany)
2,4-D	Sigma-Aldrich (Steinheim, Germany)
DTT (Dithiothreitol)	SERVA (Heidelberg, Germany)
Extraction buffer A	Sigma-Aldrich (Steinheim, Germany)
EGTA	Carl Roth (Karlsruhe, Germany)
EDTA	Carl Roth (Karlsruhe, Germany)
Flg22	GeneScript (Netherlands)
Formic acid	Merck (Darmstadt, Germany)
Gamborg B5 Medium	Duchefa Biochemie (Haarlem, Netherlands)
Glycine	Carl Roth (Karlsruhe, Germany)
Iodoacetamide	SERVA (Heidelberg, Germany
Lysonase	Novagen (Billerica, MA, USA)
Ladder blue eye prestained protein marker	Jena biosciences (Jena, Germany)
Methanol LC-MS chromasolv	Honeywell, Riedel-de Haën
Milli Q water	Milli Q device (Millipore)
Milk powder blotting grade	Carl Roth (Karlsruhe, Germany)
Macerozyme	SERVA (Heidelberg, Germany)
Nitrocellulose (parablot NCL)	Macherey-Nagel (Düren, Germany)
Nlp20	GeneScript (Leiden, Netherlands)
NIB 4x	Sigma-Aldrich kit (CELLYTPN1-1KT)
Page ruler plus, prestained protein ladder	Thermo Fisher (Rockford, IL, USA)
Potassium chloride	Merck (Darmstadt, Germany)
Protease inhibitor cocktail	Sigma-Aldrich (Steinheim, Germany)
rLys-C MS grade	Promega (Maidson, USA)
SDS pellets	Carl Roth (Karlsruhe, Germany)
Sodium deoxycholate	PanReac-Applichem (Darmstadt, Germany)
Sucrose	Sigma-Aldrich (Steinheim, Germany)
Sodium Chloride	Carl Roth (Karlsruhe, Germany)
Sodium carbonate	Carl Roth (Karlsruhe, Germany)
Tween 20	PanReac-Applichem (Darmstadt, Germany)
TEMED	PanReac-Applichem (Darmstadt, Germany)
Tris	Carl Roth (Karlsruhe, Germany)
Tris-HCL	Carl Roth (Karlsruhe, Germany)
TFA (trifluoroacetic acid)	Carl Roth (Karlsruhe, Germany)
Trypsin	Promega (Maidson, USA)
Triton x-100	Sigma-Aldrich kit (CELLYTPN1-1KT)
4′,6-Diamidine-2′-Phenylindole dihydrochloride (DAPI)	Invitrogen, Ltd. (Paisley, UK)
Urea	SERVA (Heidelberg, Germany)
Water LC-MS ultra	Honeywell, Riedel-de Haën

Software, databases and bioinformatics tools

Table 2. List of programs, software and databases

Program, softwares and databases
DAVID bioinformatics resources 6.8 https://david.ncifcrf.gov/
TAIR https://www.arabidopsis.org/
SUBA4 https://version4legacy.suba.live/
LOCALIZER 1.0.4 https://localizer.csiro.au/
STRING database https://string-db.org/
MaxQuant software https://www.maxquant.org/
Proteome Discoverer 2.1.1.21
Perseus software v.1.6.6.0 https://maxquant.net/perseus/
ImageJ https://imagej.nih.gov/ij/
ZEN (Zeiss)

Buffers & solutions

Blotting buffers

Table 3. List of blotting buffers

Name	Constituents	Concentration
Transfer buffer	10x Blot buffer Methanol	20 mL 40 mL Fill up to 200 mL with water
10x Blot buffer	Tris Glycine 10 % (w/v) SDS	15.14 g 15.01 g 10 mL Fill up to 1000 mL with water
10x TBS	Tris-HCl Sodium chloride	6.05 g 43.8 g Adjust pH 7-8, fill up to 500 mL with water
1x TBST	10x TBS Tween 20	50 mL 500 µL Fill up to 500 mL with water

SDS-PAGE buffers and solutions

Table 4. List of buffers and solutions for SDS-PAGE

Name	Constituents	Concentration
10x SDS Running buffer (Dilution to 1x was done)	Tris Glycine SDS	250 mM 2.5 M 1 % (w/v)
Stacking buffer	Tris-HCl pH 6.8 SDS	1 M 10 % (w/v)
Separation buffer	Tris-HCl pH 8.8 SDS	1.5 M 10 % (w/v)
5x Sample buffer	B- Mercaptoethanol Tris-HCl pH 6.8 SDS Glycerol Bromophenol blue	5% (v/v) 250 mM 10 % (w/v) 50% (w/v) 0.5 %

General buffers and solutions

Table 5. List of General buffers and solutions

Name	Constituents	Concentration
Protoplast enzyme solution	Macerozyme Cellulase Calcium chloride	0.2 % 0.67 % 0.24 M
B5 sucrose solution	Gamborg B5 medium 2,4-D Sucrose	0.32 % 1 mg/L 0.28 M pH 5.5
NIBA	NIB 4x (nuclei isolation buffer) DTT Protease inhibitor cocktail	25% v/v 1 mM 1 %
Different urea buffers, reducing buffers and alkylating buffers	Mentioned in details in each experiment	
Extraction buffer B	Tris base Sodium chloride SDS Sodium deoxycholate EGTA DTT Protease inhibitor cocktail	50 mM 150 mM 1 % 0.5 % 1 mM 1 mM 1.5 % pH to 7.5
Washing buffers	Potassium chloride Sodium carbonate Urea solution - Urea	1 M 100 mM 2 M

	- Tris	10 mM, pH 8
Extraction buffer (EB)	Tris base EDTA DTT SDS Protease inhibitor cocktail	50 mM 10 mM 10 mM 4 % 1.5 % pH to 8

SDS-PAGE gel preparation

Table 6. SDS-PAGE preparation

Component	Stacking gel	Separation gel (10%)
30% Acrylamide mix	2.6 mL	10 mL
1 M Tris-HCl pH 6.8	2 mL	-----------------------------------
1.5 M Tris-HCl pH 8.8	--------------------------	7.5 mL
10% SDS	0.16 mL	0.3 mL
Water	11 mL	11.9 mL
Temed	0.016 mL	0.012 mL
10 % APS	0.16 mL	0.3 mL

Devices and instruments

Table 7. List of devices and instruments

Name	Supplier
Autoclave	HMC
Autoclave, steriltechnik AG	Steriltechnik AG
Axioplan2 imaging fluorescence microscope	Carl Zeiss
Bench HERAsafe	Thermo
Beckman coulter spectrophotometer DU 800	Beckman Coulter
Balance, Sartorius cpA64	Sartorius
Centrifuge 5415D Eppendorf	Eppendorf
Centrifuge 5810R Eppendorf	Eppendorf
Confocal laser scanning microscopy LSM 900	ZEISS
Canon Scanner Lide 700 F	Canon
Developing machine for western blot, optimax 2010	Protec
Easy spray analytical column (ES803A, ES903)	Thermo scientific
Easy nLC-1000	Thermo Scientific
Fusion solo (s) chemiluminescence	Vilber
Grinder and homogenizer (percellys 24)	Bertin Technologies
Mag Rack 6	GE Healthcare, life sciences (cytiva)
ms major science electrophoresis device	ms major science
Nanodrop 800	Thermo scientific
pH Meter Mettler Toledo five easy plus	Mettler-Toledo AG

qPCR maschine Mx 3000p, stratagene	Agilent
PCR thermocycler T100	BioRad Laboratories, Inc
Phytocabinet	Percival
Q-Exactive plus Mass spectrometer	Thermo Scientific
Rotor wheel (stuart rotor SB2)	VWR
Rotor wheel (bigger eppis)	Rottberg
Shaker inforos AG, CH-4103	Inforos, bottmingen
Shaker Flatbed HS 250	IKA labortechnik
Stirrer, Ikamag REO	DREHTHAL electronics
Scanner, CanoScan LiDE 300	Canon
TransBlot SD	BioRad Laboratories, Inc
Thermomixer 5436	eppendorf
Thermomixer comfort	eppendorf
Termomixer C	eppendorf
Ultrasonic sonicator (SONOREX)	BANDELIN
Vortex GeNie 2	Bender and Hobein (Switzerland)
Vacuum concentrator Rvc2-25 CD plus	Christ (Fisher scientific)

2.2 Methods

2.2.1 Investigation of nuclear proteome using isolated nuclei from an *Arabidopsis thaliana* protoplast

(Most of the parts of the section 2.2.1 (which is my own writing) were taken as it is from my published manuscript in Frontiers in plant science. AYASH, M., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. 2021. LC-MS Based Draft Map of the Arabidopsis thaliana Nuclear Proteome and Protein Import in Pattern Triggered Immunity. Frontiers in plant science, 12, 744103-744103) licensed under CC BY 4.0.

2.2.1.1 *Arabidopsis thaliana* cell culture

A. thaliana medium consisted of 4.4 g of gamborg B5, 30 g sucrose, 5 mL of 2, 4-D (50 mg in 250 mL water) , adjusted to pH 5.9 to a final volume of 1000. The cells were grown under continuous shaking at 120 rpm at 22°C in the darkness. The cell culture was sub cultured every week or every 5 days when an experiment was planned.

2.2.1.2 Preparation of the protoplasts

Thirty mL of 5 days-old *A. thaliana* cultured cells grown in the dark were centrifuged at 805 g for 5 min at RT. The pellets were resuspended in 30 mL of 0.24 M $CaCl_2$. Then 15 mL of this suspension, 20 mL of 0.24 M $CaCl_2$ and 15 mL of the enzyme solution (0.2 % (w/v) macerozyme, 0.67 % cellulose and 0.24 M $CaCl_2$) were transferred to a Petri dish. The Petri dish was incubated at RT overnight for around 18 h with shaking at 45 rpm. The content of the Petri dish was centrifuged at 290 g for 5 min at RT and the pellets were resuspended in 30 mL

of 0.24 M CaCl$_2$. The centrifugation step was repeated and, the pellets were resuspended in 14 mL of B5 sucrose solution (0.32 % (w/v) gamborg B5 medium, 1 mg/L 2, 4-D and 0.28 M sucrose at pH 5.5). The final suspension was centrifuged at 130 g for 5 min at RT and was left for 5-10 min ar RT. The floating protoplasts were collected from the top layer and were centrifuged again same as before and then final floating protoplasts were collected. When needed, 10 µL sample from protoplast preparation was collected for microscopy. Protoplast samples were supplemented with flg22 and nlp20 respectively to a concentration of 1 µM in solution and incubated for 16 h at 18 °C. Control samples were untreated and incubated similarly. These experiments were performed three times independently (three biological replicates for each condition).

2.2.1.3 Preparation of nuclear and cellular fractions

4 mL of protoplasts were mixed with 9 mL of NIBA (25 % v/v NIB 4x (nuclei isolation buffer), 1mM DTT and 1 % (v/v) protease inhibitor) in a falcon tube and kept on ice for 10 min. Triton X-100 was added to an in solution concentration of 0.1% and the suspension was gently mixed for 5 min. Three consecutive centrifugation steps were done each at 1000 g for 15 min at 4°C. After the first 2 steps the pellets were resuspended in 4 mL NIBA containing 0.1 % triton X-100. The supernatants were combined and retained as the cellular fraction. Then the pellets were resuspended in 4 mL NIBA without Triton (washing step) and centrifuged as before. After the third step, the pellets were resuspended in 300 µL extraction buffer and transferred to a 1.5 mL Eppendorf tube (nuclear fraction, NF). NIBA 4x and extraction buffer were taken from a commercially available nuclear isolation kit (CELLYTPN1-1KT for plants, SIGMA).

2.2.1.4 DNA-staining and microscopy

100 µL of 5 µg/mL DAPI were added to 10 µL of the nuclear fraction and kept in darkness for 15 min. Then, 10 µL of this solution were used for microscopy. A fluorescence microscope (Axioplan2 imaging, Carl Zeiss) with a DAPI filter was used to visualize DAPI stained nuclei. Fluorescence was excited at about 358 nm and emission was recorded at 463 nm.

2.2.1.5 Extraction of cellular proteins

5 mL of CF was mixed with 45 mL of 100 mM ammonium acetate in methanol. The mixture was kept at -20°C overnight and then three centrifugation steps were done with a swinging bucket

rotor centrifuge at 3200 g for 15 min at 4°C. The pellets from the first two centrifugation steps were washed with 3 mL of 20% 50 mM ammonium bicarbonate and 80% acetone and the final pellets were left to dry at RT. The dried pellets were solubilized by vortexing and sonication in 450 µL urea buffer (8 M urea and 50 mM Tris) and constituted cellular proteins.

2.2.1.6 Extraction of nuclear proteins

The nuclear fraction in total was mixed to 200 µL of extraction buffer (containing 1% (v/v) protease inhibitor cocktail, Sigma P9599). Then the sample was mixed for 30 min at 1800 rpm in the RT and then followed with sonication step in an ultrasonicator for 10 min. Finally, the sample was centrifuged in a fixed rotor angle centrifuge for 10 min at 12000 g in the RT and the supernatant was collected containing the nuclear proteins.

2.2.1.7 Western blot analysis

Five µg of the protein extracts were separated into one gradient SDS-PAGE (20-4%, Serva).The proteins were transferred to a nitrocellulose membrane using a wet blot technique (Protran, GE Healthcare). The membranes were blocked with 3% (w/v) fat-free dry milk (BioRad) in TBS (50 mM Tris-HCl pH7.5, 150 mM NaCl). The blocked membranes were incubated with polyclonal anti-histone H3 antibody (Agrisera, AS10710A) and a secondary anti-rabbit antibody coupled to HRP (AS09602). Detection was performed with SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo) and the signal was recorded with the Fusion Solo S Chemiluminescence Imaging System (VWR) using a 16-bit CCD camera.

2.2.1.8 In-solution digestion of proteins using trypsin

The protein samples were reduced by addition of DTT solution (29.9 µg/µL). Then, the samples were kept at 22°C for 1 h shaking at 450 rpm. Samples were alkylated by the addition of iodoacetamide solution (35.9 µg/µL) and kept at 22°C for 1 h shaking at 450 rpm in darkness. Again, the reducing solution was added to samples and was kept at 22°C for 1 h with shaking. 50 mM ammonium bicarbonate pH 8.5 was added to each sample. Trypsin (0.2 µg/µL) was added to a ratio of 1:50. Protein digestion was allowed to proceed overnight at 37°C shaking at 750 rpm. Then the samples were dried in a vacuum concentrator.

2.2.1.9 STAGE-Tip C18 peptide desalting (Stop-and-Go Extraction)

In house produced C18-STAGE-Tips and 1.5 mL Eppendorf tubes were used in the desalting of the peptides. Firstly, the tips were conditioned with 100 µL (80% (ACN) acetonitrile, 0.1% (FA) formic acid) by centrifugation for 2 min at 1500 g at RT. Secondly, two equilibration steps were done with 100 µL 0.1% FA by centrifugation for 2 min at 1500 g at RT. The dried peptides were dissolved in 200 µL 0.1% formic acid and were added to the equilibrated tips and were centrifuged for two times each for 2 min at 1500 g at RT. All flow throughs were discarded. Two washing steps were done with 100 µL 0.1 % FA and centrifuged similarly. New Eppendorf tubes were used and the peptides were eluted by adding 50 µL of (80% ACN, 0.1% FA) and centrifugation for 1 min at 1500 g at RT. The elution step was repeated and the combined eluate were dried in a vacuum concentrator.

2.2.1.10 Liquid chromatography and mass spectrometry (LC-MS/MS)

The dried peptides were dissolved in 10 µL of (5% ACN, 0.1% TFA). The samples were analyzed on a Q Exactive Plus mass spectrometer equipped with an EASY nanoLC-1000 liquid chromatography system (both from Thermo Fisher Scientific). A flow rate of 250 nL/min was used. Peptides were separated using an analytical column ES803A (ThermoFisher) and a gradient increasing from 5% to 40% of solvent B (ACN in 0.1% FA) in 540 min followed by 13 min of isocratic flow at 80% of solvent B (for cellular proteins). On the other hand, the nuclear proteins peptide samples were separated using a gradient inclining from 5% to 35% of solvent B (ACN in 0.1% FA) in 450 min followed by 20 min of incline to 80% solvent B and finally fixed at 80% solvent B for 70 min. The spray voltage was 1.9 KV and the capillary temperature was 275°C.

A Data-Dependent Acquisition (DDA) scan strategy was used, where one MS full scan was done and then up to 10 MS2 scans of product ions from the 10 most abundant precursor ions. The MS full scan parameters were acquired at: AGC target 3E+06, resolution 70,000 and max injection time (IT) 100 ms. The MS2 parameters were executed with: resolution 17,500, Max IT 50 ms, dynamic exclusion duration 40 s, AGC target 5E+04 and isolation window 1.6 m/z.

2.2.1.11 Identification and quantification of peptides and proteins

Peptide and by inference protein identification was done by matching the MS raw data with *in silico* generated peptide ion *m/z* and MS2 spectral peak lists. The TAIR10 protein database supplemented with common contaminants (14486974 residues, 35394 sequences) was searched using the Mascot search engine V2.5.1 coupled to the Proteome Discoverer 2.1.1.21 (Thermo Fisher Scientific). The enzyme specificity was set to trypsin with tolerance of 2 missed cleavages. Ion m/z error tolerance was set to 5 ppm and 0.02 Da for precursor and fragment ions respectively. Carbamidomethylation of cysteine was set as a static modification and oxidation of methionine as a variable modification. Peptide spectral match (PSM), peptide and protein level false discovery rates (FDR) were determined by a decoy database search. A significance threshold of α= 0.01 was used for PSM and peptide level identifications. For the protein level: α of 0.05 was tolerated. The PSM count was used as protein abundance quantitative index (PQI). Protein grouping was inferred based on the principal of parsimony and only master proteins (protein group member that best explains the set of peptides used for inference) were retained. In the case of duplicate gene models producing individual master proteins, the first gene model was retained (this was the case in less than 1% of master proteins).

2.2.1.12 Computer-aided data analysis

Gene ontology analysis of the curated nuclear proteome was performed using the DAVID Bioinformatics resources 6.8 (Huang et al., 2009a, Huang et al., 2009b) using default parameters. *A. thaliana* was used as background and TAIR_ID was used as identifier. Functional annotation chart was created with threshold of count: 2 and ease: 0.1. Proteins annotated to the nucleus, with GOTERM_CC_DIRECT, were further clustered using high classification stringency. Subcellular location of the curated nuclear proteome was checked with SUBA4 (Hooper et al., 2017) using experimental locations inferred by fluorescent protein (FP) or MS/MS studies (retrieval from SUBA4 was done in January 2020 and rechecked in February 2021). LOCALIZER 1.0.4 (Sperschneider et al., 2017) was used to predict organelle subcellular localization by searching for targeting sequences such as NLS in protein primary structure and by predicting transit peptides. To further evaluate the possible biological role of the proteins in significant functional categories, their AGI codes were used to query the STRING database (Szklarczyk et al., 2019) for physical interaction setting the stringency to highest confidence

interactions, which we have shown to be true positive previously (Hoehenwarter et al., 2013), using experiments, databases, co-occurrence and co-expression as interaction sources and showing only interactions between proteins in the input set.

2.2.1.13 Collective data analysis

Mean PSM value of each protein in all measurements of the nuclear fraction (μNp_n) and the cellular fraction (μCp_n) were calculated. These values were used to formulate two scores, the nuclear and cellular enrichment scores (Npf_n and Cpf_n) that express the ratio of the abundance of protein (n) in the nuclear and the cellular fraction respectively (equation 1 and 2).

(1) $Npfn = \dfrac{\mu Np_n}{\mu Np_n + \mu Cp_n}$,   (2) $Cpfn = \dfrac{\mu Cp_n}{\mu Cp_n + \mu Np_n}$

2.2.1.14 Statistical data analysis

The matrix of curated nuclear proteins PQI (PSM) values of all samples was imported to Perseus software v.1.6.6.0 (Tyanova et al., 2016). The PQI values were grouped into 3 groups (control, flg22 and nlp20). Proteins that did not have a value in at least 5 of the 6 measurements of at least one group were discarded. The individual measurements (columns) were unit vectors normalized. Multiple sample test (ANOVA) was performed for the 3 groups in order to assess the significance of changes in abundance between conditions using permutation-based FDR multiples testing correction with an FDR significance threshold α of 0.05 and 250 permutations. Post hoc test (FDR= 0.05) was performed to identify the significant group pairs. Proteins with statistically significant changes in their abundance were kept and their values Z-score transformed. Hierarchical clustering was performed using Pearson correlation as distance measure for row clustering and Spearman correlation for columns.

2.2.1.15 2D-Quant protein concentration determination

All materials and reagents were used from the 2D-Quant kit. Working color solution was prepared by mixing 100 part of color reagent A and 1 part of color reagent B. BSA standard dilution series of 0, 10, 20, 30, 40, 50 µg were prepared from BSA stock solution (2 mg/mL). Each of them was added to 2 mL Eppendorf tube. 5 µL of protein extracts were added to 2 mL tubes (in duplicates). 500 µL of precipitant was added to each tube and was agitated briefly.

Then, 500 µL of co-precipitant was added to each tube and agitated briefly. Afterwards, centrifugation at 10000 g was done for 5 min at RT and supernatants were discarded. Protein pellets inside the tube were left for 2 min to dry and then 100 µL of copper solution and 400 µL of milli-Q water were added. The pellets were dissolved by agitation and 1 mL of working color solution was added and mixed. The tubes were left for 15 min and absorbance were measured at 480 nm with spectrophotometer using milli-Q water as blank.

2.2.2 Investigation of the nuclear proteome of *Arabidopsis thaliana* rosette leaves using proximity-dependent labelling (TurboID)

The procedures under 2.2.2.2, 2.2.2.5, 2.2.2.6, 2.2.2.7 were in house developed, established and optimized based on published protocols (Mair et al., 2019, Branon et al., 2018).

2.2.2.1 *Arabidopsis thaliana* growth conditions

A. thaliana seeds were initially sowed on soil in pots and left to grow in the phytocabinet at short day conditions (8 h day time) as follows: (1) day temperature of 22°C, day humidity at 60 % and 130 UML light intensity (2) night temperature at 20°C, night humidity at 60 % and 0 UML light intensity. After 2 weeks the young plants were transferred individually to separate pots containing soil. The plants were left to grow under the same conditions as before for another 6-7 weeks.

2.2.2.2 Biotin treatment

Adult *A. thaliana* rosette leaves (intact rosette) were detached fast with a sharp cutter and rinsed quickly with water. Rosette of each plant were added to a 1 L beaker covered with 250 mL of 50 µM biotin solution (in water) and were kept submerged for 4 h at 22°C. Afterwards, each of them was rinsed with 250 mL ice cold water followed by washing with 1 L ice cold water 3 times each for 6 min. Finally, each of them were left shortly to dry and were frozen in liquid nitrogen and stored at - 80°C.

2.2.2.3 Flg22 and cycloheximide treatments

Flg22 treatment

Leaves were infiltrated with 2 µM flg22 in solution and left for 1 h in the phytocabinet. Then the rosette leaves (intact rosette) were detached fast with a sharp cutter and treated as above under 2.2.2.2.

Cycloheximide treatment

Leaves were infiltrated with 100 µM cycloheximide solution (either alone or with flg22) and were left for 1 h in the phytocabinet. Then the rosette leaves (intact rosette) were detached fast with a sharp cutter and rinsed quickly with water. Rosette leaves of each plant were added to a 1 L beaker covered with 250 mL of (50 µM biotin solution (in water) and 100 µM cycloheximide solution). Then they were kept submerged for 4 h at 22°C. Afterwards, each of them were rinsed with 250 mL ice cold water followed by washing with 1 L ice cold water for 3 times each for 6 min. Finally each of them were left shortly to dry and were frozen with liquid nitrogen and stored at - 80°C.

2.2.2.4   Total protein extraction

400 mg of ground plant material were resuspended with 1.2 mL of extraction buffer B at pH 7.5 (50 mM Tris base, 150 mM sodium chloride, 1 % (w/v) SDS, 0.5 % (w/v) sodium deoxycholate, 1mM EGTA, 1 mM DTT and 1.5 % (v/v) protease inhibitor cocktail). Samples were vortexed and then mixed for 10 min at 95°C then for 20 min at 22°C. 0.6 µL of lysonase was added and then the mixture was mixed for 15 min at 22°C. Afterwards, it was sonicated for 5 min in cold water and then centrifuged at 16000 g for 10 min at 10°C. The supernatant was transferred to a new Eppendorf tubes and was centrifuged at 20000 g for 30 min at 10°C. The supernatant was transferred to a new tube and used immediately.

2.2.2.5   Removal of excess free biotin by PD-10 gel filtration columns

The PD-10 gel filtration columns were equilibrated five times each with 5 mL cold extraction buffer B without protease inhibitor. 2.5 mL protein extract was applied and allowed to enter the column bed completely. All flow throughs were discarded. Proteins were eluted with 3.5 mL of cold extraction buffer B without protease inhibitor.

2.2.2.6   Enrichment of biotinylated proteins on streptavidin beads

Dynabeads Myone Streptavidin C1 were first washed according to the manufacturer's protocol. The beads were vortexed for 1 min and then the desired volume was transferred to a new 2 mL Eppendorf tubes. Equal amount or at least 1 mL of extraction buffer B without protease inhibitor was added to the beads and the mixture was resuspended. The tube was placed on a magnet

for 1 min to precipitate beads and the supernatant was discarded. The tube was removed from magnet and the beads were resuspended again in a volume of extraction buffer B equal to initial amount of beads taken from vial. The mixture was resuspended and separated on magnet again. These steps were performed for 3 times. For each sample a volume of protein extract equivalent to 16 mg of protein amount (supplemented with 1.5% (v/v) protease inhibitor) was split and applied to four 5 mL Lo-bind Eppendorf tubes each containing 100 µL of washed beads. The samples were incubated on a rotor wheel at 4°C overnight for 16 h. The following day, the beads were separated from the extract on a magnetic rack and were washed as follows: (1) two times with cold extraction buffer B (2) one time with cold 1 M potassium chloride (3) one time with cold 100 mM sodium carbonate (4) one time with 2 M urea in 10 mM tris at pH 8 at RT (5) two times with cold extraction buffer B. All washes were done with 1 mL and for 8 min with rotation on a rotor wheel. All supernatant washing solutions were discarded after separation on a magnetic rack. The beads were subsequently used without storage.

2.2.2.7   On-bead digestion of proteins

1 mL of 50 mM tris at pH7.5 was added to the beads and mixed on a rotor wheel for 8 min for two times. The beads were transferred to a new 1.5 mL Eppendorf tube and were mixed with 1 mL 2 M urea in 50 mM tris at pH 7.5 for 8 min and the beads were separated on a magnetic rack. 80 µL of trypsin buffer (50 mM tris pH 7.5, 1 M urea and 1mM DTT) and 2 µL of trypsin (0.2 µg / µL) were added to the beads and incubated for 3 h shaking at 800 rpm at 25°C. The supernatant was transferred to new tubes and the beads were washed twice each with 60 µL of trypsin buffer without trypsin. The supernatants were collected and pooled with the initial supernatant to give a final volume of 200 µL. The solution was reduced by 4 mM DTT with mixing at 450 rpm at 25°C. Then the solution was alkylated by 10 mM iodoacetamide with mixing at 450 rpm in the dark at 25°C. Finally, 2.5 µL of trypsin (0.2 µg / µL) were added and the solution was incubated at 25°C shaking at 800 rpm overnight for 15 h. Then the samples were dried in a vacuum concentrator. The desalting steps were done the same as under 2.2.1.9.

2.2.2.8   Liquid chromatography and mass spectrometry (LC-MS/MS)

The dried peptides were dissolved in 12 µL of (5% ACN, 0.1% TFA). The samples were analyzed on a Q Exactive Plus mass spectrometer equipped on-line with an EASY nanoLC-1000 liquid chromatography system (both from ThermoFisher Scientific). A flow rate of 250 nL/min was used. Peptides were separated using an analytical column ES903 (ThermoFisher)

and a gradient increasing from 5% to 40% of solvent B (ACN in 0.1% v/v FA) in 540 min followed by 13 min of isocratic flow at 80% of solvent B. The spray voltage was 1.9 KV and the capillary temperature was 275°C.

A Data-Dependent Acquisition (DDA) scan strategy was used, where one MS full scan was performed followed by up to 10 MS2 scans of product ions from the 10 most abundant precursor ions. The MS full scan parameters were acquired at: AGC target 3E+06, resolution 70,000 and max injection time (IT) 100 ms. The MS2 parameters were: resolution 17,500, Max IT 50 ms, dynamic exclusion duration 40 s, AGC target 5E+04 and isolation window 1.6 m/z.

2.2.2.9 Identification and quantification of peptides and proteins

Peptide and by inference protein identification was done by matching the MS raw data with *in silico* generated peptide ion *m/z* and MS2 spectral peak lists. The TAIR10 protein database supplemented with common contaminants (14486974 residues, 35394 sequences) was searched using MaxQuant version 2.0.1.0. All parameters on MaxQuant were left as default settings with minor modifications.The enzyme specificity was set to trypsin/p with tolerance of 2 missed cleavages. Carbamidomethylation of cysteine was set as a static modification and oxidation of methionine and N-terminal acetylation as a variable modification. Maximum number of modification per peptide was set to 5. PSM FDR and protein FDR was set to 0.01. For protein quantification unique and razor peptides were used. Data normalization was done with MaxQuant algorithm method. The LFQ (label free quantification, normalized intensity) was used as protein abundance quantitative index (PQI). LFQ minimum ratio count was set to 1 and fast LFQ was checked. Match between runs and second peptides were checked. Majority protein IDs of protein groups were used and in the case different proteins are present then only leading proteins with the highest number of identified peptides were retained. Proteins that are marked as only identified by site, reverse and potential contaminants were discarded

2.2.2.10 Computer-aided data analysis

Gene ontology analysis of the nuclear proteome was performed using DAVID Bioinformatics resources (Huang et al., 2009a, Huang et al., 2009b, Sherman et al., 2022) using default parameters. *A. thaliana* was used as background and TAIR_ID was used as identifier. Functional annotation chart was created with threshold of count: 2 and ease: 0.1. Proteins annotated to the nucleus, with GOTERM_CC_DIRECT, were further clustered. Subcellular

location of the nuclear proteome was checked with SUBA4 (Hooper et al., 2017) using experimental locations inferred by fluorescent protein (FP) or MS/MS studies. LOCALIZER 1.0.4 (Sperschneider et al., 2017) was used to predict organelle subcellular localization by searching for targeting sequences such as NLS in protein primary structure and by predicting transit peptides.

2.2.1.11 Statistical data analysis

The protein groups file was extracted from MaxQuant and was imported to Perseus software v.1.6.6.0 (Tyanova et al., 2016). The PQI (LFQ) values were grouped and log2 transformed. Proteins were filtered and only proteins with minimum 3 values in at least one group were retained. Missing values were imputed from the normal distribution with default settings (width: 0.3, down shift: 1.8 and mode: separately for each column). Multiple samples test (ANOVA) was performed for the 4 groups (flg22 (F), flg22+cycloheximide (FC), cycloheximide (C) and water treated control (W)) in order to assess the significance of changes in abundance between conditions using permutation-based FDR multiples testing correction with an FDR significance threshold α of 0.05 and 250 permutations. Post hoc test (FDR= 0.05) was performed to identify the significant group pairs. Proteins with, significant change in abundance in between flg22 (F) and water (W), were extracted and filtered to remove proteins that are absent in 3 replicates in water conditions. Then their values were Z-score transformed. Heat map was generated and hierarchical clustering was performed using Pearson correlation as a distance measure for row clustering and Spearman correlation for columns.

2.2.2.12   SDS-PAGE and western blot analysis

Samples were analyzed on SDS-PAGE to separate proteins according to molecular weight. 10% polyacrylamide gels were used. Separation was done by electrophoresis under 120 V. Samples were prepared in two ways. In the first one, frozen leaves were ground with a homogenizer using tubes and small beads and then the ground material was resuspended in 1x sample buffer (diluted from 5x sample buffer, Table 4) boiling for 5 min at 95°C. In the second one, protein extracts were mixed with 5x sample buffer (Table 4) so that the final buffer concentration was at least 1x. Then the mixture was mixed and boiled for 5 min at 95°C. Protein size was determined using the blue eye prestained protein marker PS-104 (Jena biosciences) and the page ruler plus prestained protein ladder 26619 (Thermo). After separation on SDS-

PAGE the proteins were blotted onto a nitrocellulose membrane using a current of 0.8 mA/cm^2 for 1 h. Then, the nitrocellulose paper was blocked with either 5% Milk in TBST (Table 3) or 3% BSA in TBST for 1 h. Afterwards, the membrane was incubated overnight at 4°C with either 3% Milk TBST containing the Anti-GFP polyclonal antibody or 1% BSA in TBST containing the Streptavidin-POD conjugate. On the next day, the membrane was washed with 1x TBST 6 times each for 10 min (GFP) and with 1x TBST 6 times each for 5 min followed by 4 times with 1x TBS each for 5 min (Streptavidin). In the case of GFP, anti rabbit secondary antibody was added and incubated for 1 h shaking at RT. Next, a second washing step was done as before. ECL prime Amersham western blotting detection reagent and Pierce ECL plus western blotting substrate were used for visualization of protein bands. Finally the membrane was fixed on a metal cassette and a film was developed using the developing machine for western blotting (optimax 2010).

2.2.2.13 Microscopy

For fluorescence microscopy, leaves were cut into small piece and added to Eppendorf tubes filled with water and suction was applied with a syringe to remove the air from the tissues. Then the leaf lower part was examined with an Axioplan2 imaging fluorescence microscope using the YFP filter. For confocal microscopy, the plant leaf was cut into discs and examined with confocal laser scanning microscopy LSM 900 using a laser at 488 nm with optical selection at 1.1 µM and with 410 – 545 nm for YFP detection and 635 – 640 nm for chlorophyll detection.

2.2.2.14   Confirmatory experiment to check for the effectiveness of cycloheximide treatment

Part of protein extracts prepared in 2.2.2.4 were digested, reduced and alkylated with the FASP protocol which was developed based on previously published papers (Song et al., 2018, Su et al., 2018, Wiśniewski et al., 2009). Amicon ultra 30K filter were added to 5% (v/v) tween 20 shaking at 60 rpm overnight. On the next day, the filters were rinsed with milli Q water and then washed for 30 min. The washing step was repeated twice exchanging the milli Q water. Protein extract containing 100 µg of protein was added to the filter and the volume was filled up to 200 µL with urea solution (50 mM tris base and 8M urea, pH 8). Centrifugation was done at 16000 g for 10 min and the flow-through was discarded. Then, 200 µL of urea solution were added and centrifugation was done at 16000 g for 10 min. The flow-through was discarded. This step was repeated twice. Then, 100 µL of reducing solution (50 mM tris base, 8M urea and 100 mM DTT, pH 8) were added and the sample was mixed at 600 rpm for 1 h at 22°C and centrifugation was

done as before. Flow-through was discarded. 100 μL of alkylating solution (50 mM tris base, 8M urea and 100 mM iodoacetamide, pH 8) were added and the sample was mixed at 600 rpm for 1 h in the dark at 22°C and centrifuged as before. The flow-through was discarded. 200 μL of urea solution were added and the sample was centrifuged as before. The flow-through was discarded. This step was repeated twice. 100 μL of digestion solution (50 mM ammonium bicarbonate, pH 8) were added and the centrifuged as above. The flow-through was discarded. This step was repeated twice. Afterwards, 50 μL of digestion solution were added and then 10 μL of trypsin (0.2 μg / μL) were added. The mixture was left mixing at 600 rpm overnight at 37°C. Finally, the filter was transferred to a new collection tube and was centrifuged as before. The filtrate was retained. Then, 40 μL of digestion solution were added and centrifuged. This step was repeated twice and all filtrates was combined with the first one representing the digested peptides. The peptides were dried in vacuum concentrator. The desalting steps were done the same as under 2.2.1.9.

The dried peptides were dissolved in 25 μL of (5% ACN, 0.1% TFA). The samples were analyzed on a Q Exactive Plus mass spectrometer on-line with an EASY nanoLC-1000 liquid chromatography system (both from ThermoFisher Scientific). A flow rate of 250 nL/min was used. Peptides were separated using an analytical column ES903 (ThermoFisher) and a gradient increasing from 5% to 40% of solvent B (ACN in 0.1% FA) in 180 min. The spray voltage was 1.9 KV and the capillary temperature was 275°C.

A Data-Dependent Acquisition (DDA) scan strategy was used, where one MS full scan was done and then up to 10 MS2 scans of product ions from the 10 most abundant precursor ions. The MS full scan parameters were acquired at: AGC target 3E+06, resolution 70,000 and max injection time (IT) 100 ms. The MS2 parameters were: resolution 17,500, max IT 200 ms, dynamic exclusion duration 50 s, AGC target 1E+05 and isolation window 1.6 m/z.

Peptide and by inference protein identification was done by matching the MS raw data with *in silico* generated peptide ion *m/z* and MS2 spectral peak lists. The TAIR10 protein database supplemented with common contaminants (14486974 residues, 35394 sequences) was searched using MaxQuant version 2.0.1.0. All parameters on MaxQuant was left as default settings with minor modifications.The enzyme specificity was set to trypsin/p with tolerance of 2 missed cleavages. Carbamidomethylation of cysteine was set as a static modification and oxidation of methionine and N-terminal acetylation as a variable modification. Maximum number of modification per peptide was set to 5. PSM and protein FDRs were set to 0.01. For protein

quantification unique and razor peptides were used. Data normalization was done by the Maxquant algorithm method. The LFQ (label free quantification, normalized intensity) was used as protein abundance quantitative index (PQI). LFQ minimum ratio count was set to 1 and fast LFQ was checked. Match between runs and second peptides were checked. Majority protein IDs of protein groups were used and in the case protein ambiguity only leading proteins with the highest number of identified peptides were retained. Proteins that are marked as only identified by site, reverse and potential contaminants were discarded.

The protein groups file was extracted from MaxQuant and was imported into the Perseus software v.1.6.6.0 (Tyanova et al., 2016). The PQI (LFQ) values were grouped and $\log_2$ transformed. Proteins were filtered and only proteins with minimum 3 values in at least one group were retained. Missing values were imputed from normal distribution with default settings (width: 0.3, down shift: 1.8 and mode: separately for each column). Multiple samples test (ANOVA) was performed for the 3 groups (flg22 (F), flg22+cycloheximide (FC) and water treated control (W)) in order to assess the significant changes in abundance between conditions using permutation-based FDR multiples testing correction with an FDR significance threshold α of 0.05 and 250 permutations. Post hoc test (FDR= 0.05) was performed to identify the significant group pairs. Proteins with statistically significant changes in their abundance were kept and their values Z-score transformed. Proteins with significant increase in abundance in flg22 (F) against water (W) and not in flg22+cycloheximide (FC) against water (W), were extracted and used to generate a heat map.

2.2.2.15 RNA extraction using SV total RNA isolation system kit

Protocol used was based on (Kobs, 1998). 175 µL of RNA lysis buffer were added to 30 mg of ground plant material. Then, 350 µL of RNA dilution buffer were added and mixing was done by inversion followed by centrifugation at 16000 g for 10 min. 200 µL of 95 % ethanol were added to the lysate and mixed for 3 to 4 times, transferred to a spin column device and centrifuged at 13000 g for one min. Liquid from the collection tube was discarded and 600 µL of RNA wash solution were added to the spin column device and centrifugation was done at 13000 g for one min. Collection tube content was discarded as before. 50 µL of DNASE incubation mix (40 µL yellow buffer, 5 µL 0.09 M $MnCl_2$ and 5 µL DNASE I enzyme) were added to the spin basket and incubated at 22°C for 15 min. Then, 200 µL DNASE stop solution were added and centrifuged at 13000 g for 1 min. Then, 600 µL of RNA wash solution were added and centrifuged as before. The collection tube was emptied and then 250 µL RNA wash solution

were added and centrifuged for 2 min. The spin basket was removed from collection tube and placed onto the elution tube. Then, 100 μL nuclease free water were added and centrifuged at 13000 g for 1 min. Finally, the concentration of the RNA was measured on a Nanodrop device using standard procedures. The elution tube containing the RNA was stored at -80°C.

2.2.2.16 cDNA synthesis

0.5 μg of RNA were added to 1.1 μL of 10x buffer, 1 μL of DNASE I was added and the volume was completed to 11 μL with nuclease free water. The mixture was incubated for 30 min at 37°C and then 1 μL of 25 mM EDTA, pH 8 was added and followed by incubation for 10 min at 65°C. Master mix was prepared as follows: 4 μL 5x buffer, 1 μL 100 μM oligo (dT) 18 primer, 2 μL 10 mM dNTPs, 1 μL ribolock and 1 μL revertaid reverse transcriptase were mixed. 9 μL of master mix was added to 12 μL of mixture prepared as decribed above. Incubation was done in a PCR machine with these parameters: 5 min at 37°C, 60 min at 42°C and 10 min at 70°C. Finally the cDNA solution was diluted 1:10 with nuclease free water and stored at -20°C.
All materials were taken from DNase I, RNase-free kit and revert aid first strand cDNA synthesis kit from Thermo Fisher.

2.2.2.17 qPCR analysis

Master mix was prepared as follows: 0.18 μL primer forward, 0.18 μL primer reverse, 10 μL 2x evagreen mix and 6.64 μL autoclaved milli-Q water. 3 μL of synthesized cDNA were mixed with 17 μL master mix. Then qPCR analysis was done in a qPCR machine with the following parameters: incubation at 95°C for 15 min followed by a total of 40 cycles comprising 15 seconds denaturation at 95°C followed by 40 seconds annealing and elongation at 65 °C.
All materials were taken from 5x QPCR Mix EvaGreen® (Rox) kit (Bio&SELL GmbH).

3. Results

The model plant *A. thaliana* was used to investigate the plant nuclear proteome in pattern triggered immunity. LC-MS-based proteomics methods were used employing two different approaches. The first one is based on nuclei isolation from protoplasts of cell culture and the second on proximity dependent labelling of nuclear proteins in adult rosette leaves.

3.1 Investigation of nuclear proteome using isolated nuclei from an *Arabidopsis thaliana* protoplast.

(Most of the parts of the section 3.1 (which is my own work and writing) were taken as it is from my published manuscript in Frontiers in Plant Science. AYASH, M., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. 2021. LC-MS Based Draft Map of the Arabidopsis thaliana Nuclear Proteome and Protein Import in Pattern Triggered Immunity. Frontiers in plant science, 12, 744103-744103) licensed under CC BY 4.0.

In this approach, we set out to produce a high-quality draft catalog of the *A. thaliana* nuclear proteome based on isolation of nuclei and mass spectrometric (MS) measurement of nuclear protein fractions. The isolation of the nucleus from the protoplasts was based on several centrifugation and dilution steps using 0.1 % Triton. The method, used in this section, was developed and optimized based on nuclear isolation kit protocol (CELLYTPN1-1KT) and in house protocols including a protocol in a previous master thesis work in our lab (Abukhalaf, 2018). The method main steps are summarized in the following scheme (Figure 8). Beyond this, we investigated the quantitative changes in protein abundance in the nuclear proteome in the three biological scenarios (control, flg22 and nlp20). To do this we chose to expose *Arabidopsis* protoplasts released from cells in culture by enzymatic digestion of cell walls to the elicitors. Additionally, we took first steps towards identifying putative candidate proteins imported into the nucleus in two related forms of PTI elicited using flg22 and nlp20 respectively on a large scale.

In order to specifically gain access to the nuclear proteome, nuclei were isolated from the protoplast incubated at 18 °C for 16 h under 3 conditions: 1 µM flg22, 1 µM nlp20 and untreated in case of control. This was repeated twice for a total of three independent experiments. The cellular suspension resulting as a product of the isolation procedure was also retained and used to prepare the cellular protein fraction comprising all proteins with the exception of those in the nucleus. The isolated nuclei were characterized by fluorescence microscopy after DAPI staining which attested to their successful isolation in an intact and round form (Figure 9). Nuclear proteins were extracted, and in-solution digested with trypsin along with the cellular proteins. The dissolved peptides were analyzed with liquid chromatography mass spectrometry (LC-MS) using a Data-Dependent Acquisition (DDA) scan strategy. In all three experiments, the MS

analysis identified 3899, 3212 and 3081 protein groups in total (set of proteins identified with a non-redundant peptide set, hence referred to as proteins) in the nuclear fraction of untreated, flg22 treated and nlp20 treated samples, respectively. Likewise, 5633, 4742 and 5636 proteins were identified in the cellular fractions of the respective samples (Appendix 1, Supplementary file 1-Tables 1-3 and 7-9) (Ayash et al., 2021). The overlap between the fractions was 2587, 2301 and 2252 proteins respectively (Figure 10 A).
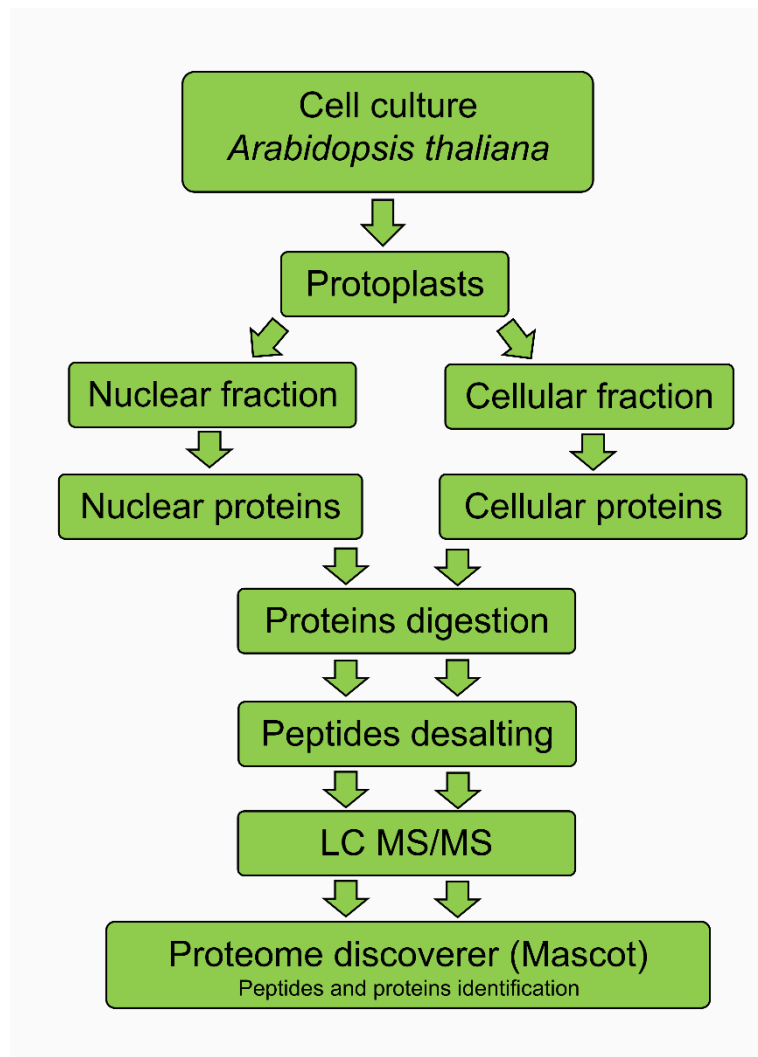
Figure 8. Summarized workflow of the nuclei isolation method. Cell cultures were grown and the protoplasts released by enzymatic digestion of cell walls. Nuclear and cellular fractions were prepared followed by extraction of nuclear and cellular proteins. Trypsin digestion of proteins was performed followed by desalting of digested peptides. The dissolved peptides were

analyzed with liquid chromatography mass spectrometry (LC-MS) using a Data-Dependent Acquisition (DDA) scan strategy and identified by Proteome Discoverer using (Mascot software).
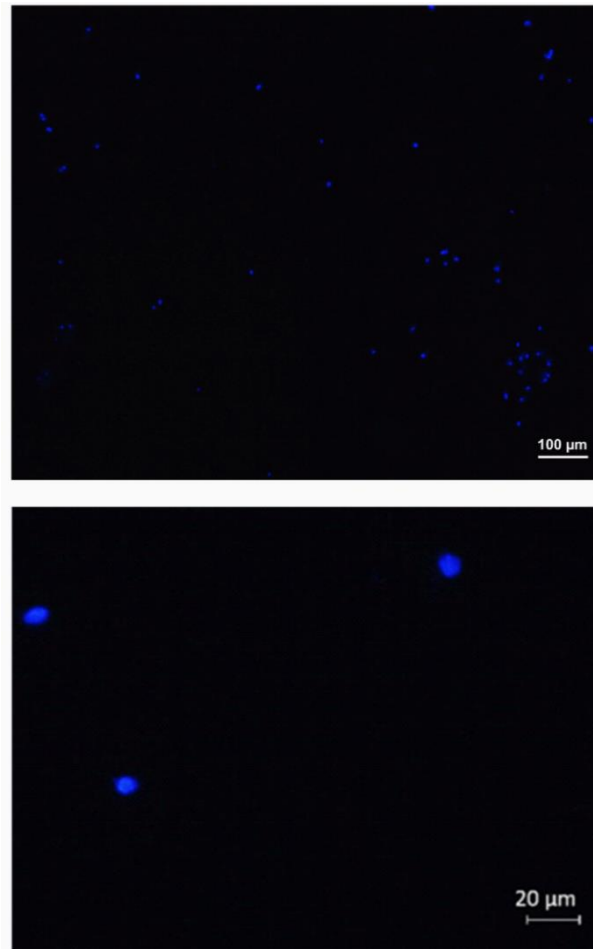
Figure 9. DAPI staining fluorescence microscopy of the nuclei isolated from protoplasts. The nuclei were isolated from dark grown *Arabidopsis* cell culture and 100 µL of 5 µg/mL DAPI were added to 10 µL of the nuclear fraction and kept in darkness for 15 min and then examined with an Axioplan2 imaging (Carl Zeiss) fluorescence microscope. Three independent experiments were done and the data shown are representative for these experiments

3.1.1 Defining the *Arabidopsis thaliana* nuclear proteome.

A central issue in all organelle isolation procedures is the purity and integrity of the preparation. Conventionally the purity of the extracted nuclear proteome is assessed by western blot of nuclear protein markers, often histones. This was done using antibody against histone H3

(Figure 10 B) but in addition an approach was devised to assess quality directly from the MS data. The fraction of each protein's abundance in the nuclear and cellular protein fractions was calculated independently by way of the acquired MS data as mentioned in the methods section under collective data analysis. This can be interpreted as an enrichment score, which in brief is the ratio of a protein's MS signal in the nuclear or cellular fraction to its total MS signal in both fractions. Thus exclusive detection in the nuclear fraction would give a nuclear enrichment score (Npf_n) of one whereas the score would be 0 if it were detected only in measurements of the cellular fraction. The score for the cellular fraction (Cpf_n) would be the inverse.

The median nuclear enrichment score (Npf_n) of all histones in the control, flg22 and nlp20 samples is shown in Figure 10 C (for each histone see appendix 1 Supplementary file 1, Tables 4-6) (Ayash et al., 2021). All samples show enrichment of histones indicating successful isolation of nuclei and extraction of the nuclear proteome. In addition, Npfn values of SUN2, KAKU4, WIP3, WIT1 and MAD1/NES1, five NE / INM proteins, were equal to 1 (exclusive presence in the nuclear protein fraction) indicating nuclei were isolated with the NE largely intact. To get an impression of the extent of inevitable contamination of the experimental nuclear proteome by cellular proteins, the fraction of the abundance of nine *bona fide* cytoplasmic markers (phosphoenolpyruvate carboxylase 1, phosphoenolpyruvate carboxylase 2, phosphoenolpyruvate carboxylase 3, Actin 1-3, Actin 7, Actin 8, Actin 12, sucrose phosphate synthase 1F and sucrose phosphate synthase 2F) were used. In contrast to histones, the median nuclear enrichment score (Npf_n) of these markers was low (Figure 10 C, left). In addition, we calculated median Npf_n values for the known mitochondrial and Golgi apparatus markers, voltage dependent anion channel 1, 2 and 3 as well as isocitrate dehydrogenase 1 and subunit 2 (mitochondrion) and coatomer gamma-2 subunit (Golgi apparatus) which also were in the range of the cytoplasmic markers (Figure 10 C, right). FD-GOGAT and FNR1 and 2 which are known plastid markers were completely absent from nuclear fractions. Together these results show that the isolation of nuclei was successful and of high purity.

Regarding the nuclear proteome, the proteins shared by both nuclear and cellular fractions (Figure 10 A, intersections) may indeed be common to both and underlie some type of trafficking between nucleus and other organelles or cytoplasm or may simply be inevitable experimental contaminations of the nuclear fraction. To address this issue, the median Npf_n values of the cytoplasmic markers (0.27, 0.24 and 0.24 respectively) described above, which are known not be present in the nucleus, were used as arbitrary cut off limits to define contamination in the three biological conditions and produce a curated set of nuclear proteins.

All proteins with Npf_n values higher than this cut off limit were considered as genuinely localized in the nucleus and thus as nuclear proteins under the applied experimental conditions whereas those proteins with Npf_n values lower than the cut off limit were considered experimentally produced cellular protein contaminants and were discarded. The used cytoplasmic markers were discarded. This led to a curated set of nuclear proteins consisting of 2839, 2259 and 2096 proteins under control, flg22 and nlp20 conditions (Appendix 1, Supplementary file 1- Tables 4-6) (Ayash et al., 2021).

To further validate the nuclear proteomes, the curated protein lists were analyzed with the DAVID Bioinformatics resources 6.8 gene ontology tool (Huang et al., 2009a, Huang et al., 2009b) and LOCALIZER (Sperschneider et al., 2017), a software that predicts organelle subcellular localization by searching for targeting sequences such as NLS in protein primary structure. Also our experimentally determined proteomes were compared with previously published nuclear / sub-nuclear proteomes and nuclear localized proteins by FP (Bae et al., 2003, Pendle et al., 2005, Calikowski et al., 2003, Bigeard et al., 2014, Sakamoto and Takagi, 2013, Chaki et al., 2015, Goto et al., 2019, Palm et al., 2016, Mair et al., 2019, Tang et al., 2020, Hooper et al., 2017). As a result, 89 % of the nuclear proteins in each condition consisted of either experimentally verified nuclear proteins or proteins annotated / predicted to be localized in the nucleus (redundancy was removed), (Figure 11, left). (Appendix 1, Supplementary file 1- Tables 4-6). This underscores the high quality of the nuclear proteome preparation.
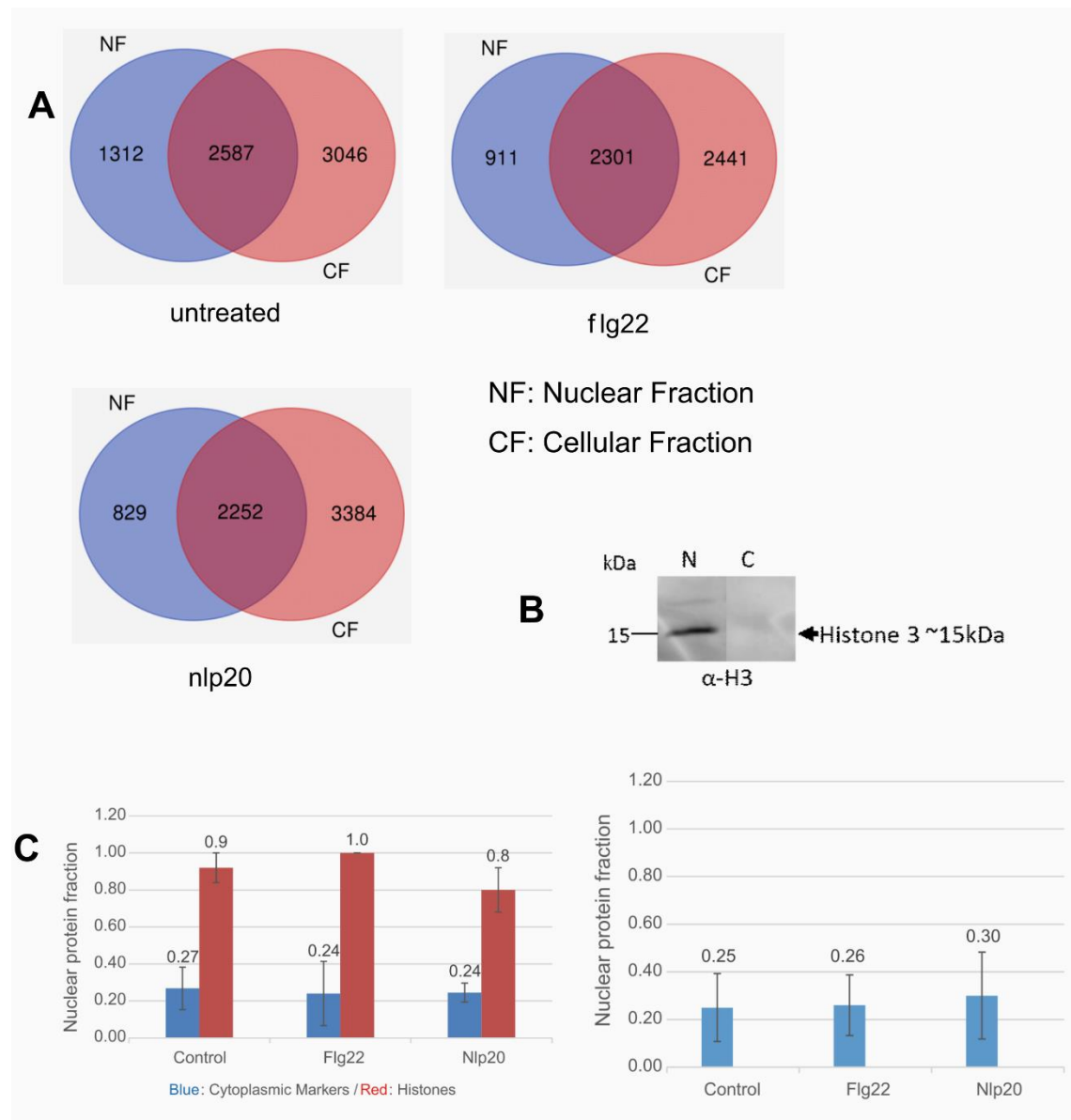
Figure 10. Defining the nuclear proteome 1 (nuclei isolation method). A. Total proteins identified in nuclear and cellular fractions in three independent experiments in untreated, flg22 and nlp20 treated cell cultures. **B**. Western blot of nuclear (N) and cellular (C) proteins with anti-Histone H3 antibody, respective MW range is shown, this W.B was done in the lab of Mareike Heilmann. **C**. (Left hand panel) Median nuclear enrichment scores (Npfn values) in three independent experiments of all identified histone proteins and of cytoplasmic markers (two cytoplasmic markers were absent in shared proteins in flg22 and nlp20 conditions and were not used in both conditions). Numbers above the bar are median Npfn values. Error bars denote median absolute deviation. **C**. (Right hand panel) Median Npfn values in three independent experiments of mitochondrial and Golgi markers. Numbers above the bar are median Npfn values. Error bars denote median absolute deviation.

The DAVID bioinformatics tool was used to further annotate the nuclear proteins and classify them according to their function. 1426 proteins of the proteome measured under untreated conditions were classified initially as belonging to the nucleus with a Benjamini corrected p-value of 1.3E-44. This set was then further input into DAVID. These proteins were categorized into 83 clusters and six main protein classes (Figure 11, right and appendix 1, Supplementary file 1- Table 10). The six main classes were transcription (231 protein), ATP binding & kinases (162 proteins), nucleotide binding (76 proteins), ribosomal proteins (36 proteins), WD40 (33 proteins) and translation initiation factor activity (17 proteins).

Proteins annotated as related to the process of transcription, i.e. transcription factors (TFs) and transcriptional co-activators were classified into families, as shown in (Table 8) and (Appendix 2, Supplementary table 1), a total of 258 in all three conditions; control and elicited. The top three transcription factors families pertaining to the number of proteins identified were bZIP (13 proteins), WRKY (8 proteins) and Trihelix (8 proteins). Nuclear envelope proteins were also identified. In control conditions, 16 proteins were reported to be localized in the nuclear envelope, inner and outer membrane based on FP experimental evidence in SUBA4. Moreover, 350 proteins were reported to be localized to the same locations based on FP and MS experimental evidence in SUBA4. All together, these results constitute a high-quality catalog of the nuclear proteome of *A. thaliana* cell culture under homeostasis and induced immunity.
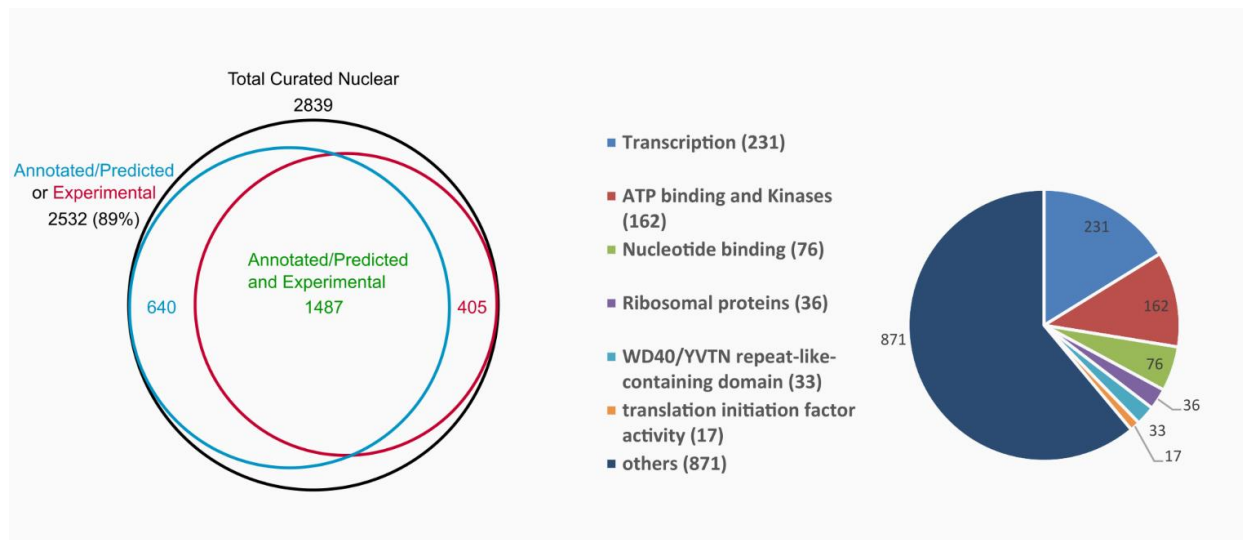
Figure 11*.* **Defining the nuclear proteome 2 (nuclei isolation method).** Left hand panel. VENN diagram showing sets of experimentally validated or annotated/predicted nuclear proteins, comprising our data set. Right hand panel. DAVID gene ontology classification of 1426 proteins of the untreated nuclear protein fraction annotated as nuclear proteins by DAVID

bioinformatics tool. Untreated condition was used as an example to generate these two figures. Three replicates were used.

Table 8. Proteins annotated to transcription process were classified into families. UP_KEYWORDS category was used in the David bioinformatics resources annotation**.** Control and elicited conditions were used to generate this list.

Families (Transcription factors)	Number of proteins	Families (Transcriptional regulators)	Number of proteins
bZIP	13	ARID	3
Trihelix	8	PHD	3
WRKY	8	LUG	2
bHLH	7	GNAT	2
global	7	SWI/SNF-SWI3	1
CCAAT	7	SWI/SNF-BAF60b	1
HB	6	HMG	1
MYB related	5	Jumonji	1
NAC	5	MBF1	1
GeBP	5		
MYB	4		
C2H2	4		
ARF	4		
Alfin-like	4		
orphans	3		
AP2	2		
BES1	2		
E2F	2		
TCP	2		
HSF	2		
Coactivator p15	2		
SNF2	1		
ABI3VP1	1		
C2C2-Dof	1		
TAZ	1		
AP2-EREBP	1		
GRAS	1		
BPC	1		
SAP	1		
BSD	1		

3.1.2 LC-MS-based protein import into the nucleus under flg22 and nlp20 stimulus

The curated nuclear proteins, identified under control, flg22 and nlp20 conditions, were compared as shown in (Figure 12 A). 1525 proteins were common to all three conditions, 269 and 223 were specific to flg22 and nlp20, respectively. This means that in these experiments considering our detection limits, these proteins appeared in the nucleus after elicitation of immunity with one of the two elicitors. These proteins could originate from the cytosol or other organelles and be imported upon PAMP perception or newly synthesized and then imported. In order to investigate this further, the cellular proteins, measured under the non-elicited control conditions, were checked for the presence of these specific proteins. 157 out of the 269 proteins appearing in the nucleus after flg22 exposure and 73 out of the 223 proteins appearing after nlp20 exposure were also measured in the cellular fraction without elicitation. Secondly, the nuclear enrichment (Npfn) and cellular enrichment scores (Cpfn) of the proteins were compared between control and flg22 and nlp20 elicited samples for the two sets of putatively imported proteins (157 and 73 proteins respectively). The Cpfn decreased in both sets upon induction of immunity with either flg22 or nlp20 when compared to untreated samples. The Npfn increased proportionally in elicited samples when compared to control (Figure 12 B and C). This indicates that these sets of candidate proteins could be trafficked to the nucleus from the cytosol or some other cellular organelle upon elicitation of immunity wherein they could play some function. Both sets of proteins are shown in appendix 2, Supplementary table 2.

We were interested in the re-localization of proteins from mitochondrion to the nucleus. To investigate this further, the 157 putatively imported proteins were checked for the presence of predicted mitochondrion transit peptides by Localizer and we found 18 proteins with transit peptide predictions. Secondly, the MS raw data was researched with no enzyme specificity to identify peptides with non-tryptic N-termini generated by in vivo cleavage. If these non-tryptic cleavages demarcate the protein's N-terminus and match transit peptide cleavage sites, then it suggests transit peptide cleavage of the protein in vivo. As shown in Table 9, the initial N-terminal part of primary structure of two proteins contain identified peptides with non-tryptic N-terminal cleavage sites (no R or K before, F was found in both), indicating the peptide sequences preceding the identified peptides (1-30 in first protein and 1-25 in second) were not cleaved by trypsin. This could imply that the two proteins were identified in the nucleus in an already cleaved form without the peptide sequences 1-30 and 1-25 respectively. These two proteins are known mitochondrial proteins, the first one contains a transit peptide at position 1-

30 as investigated before (Carrie et al., 2015) and the second has a transit peptide at position 1-24, predicted by Localizer. This suggest that these two proteins may be re-localized to the nucleus in their cleaved forms.
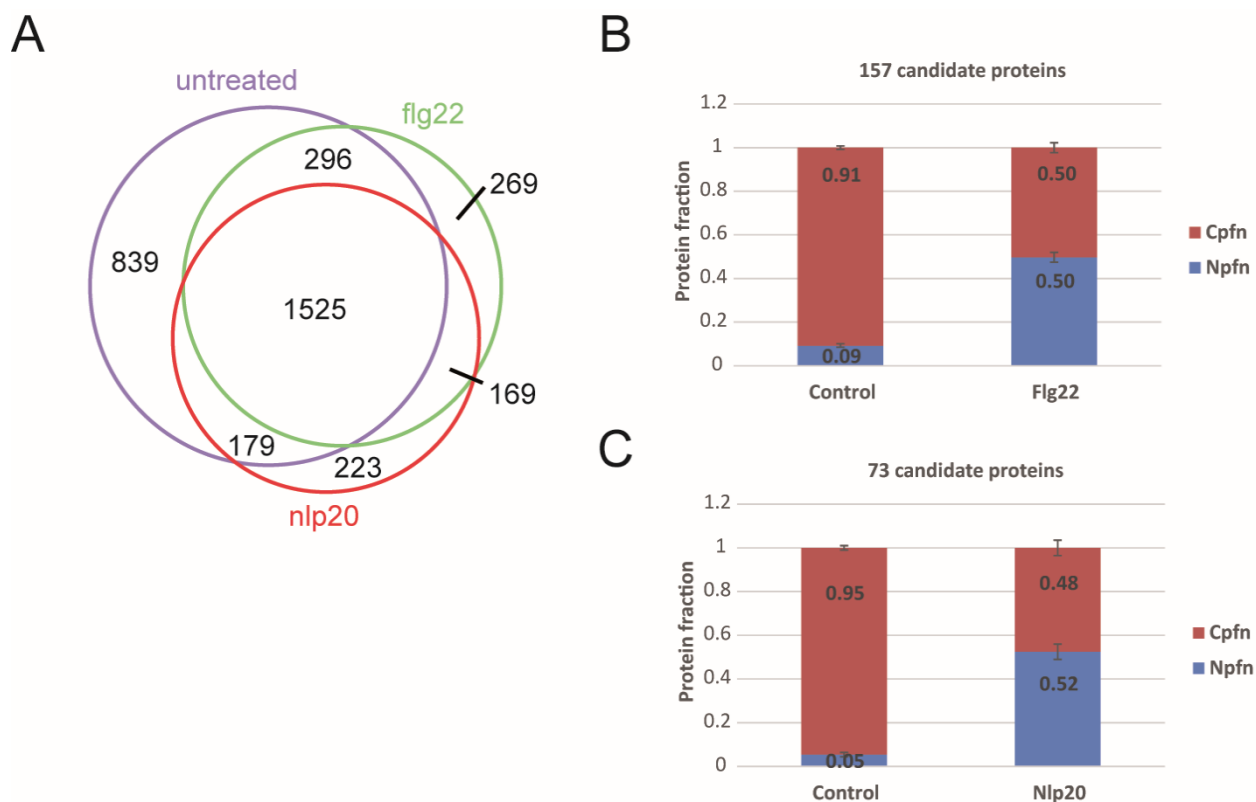
Figure 12. Protein import into the nucleus following elicitation of PTI. A. Intersections of curated nuclear protein fractions extracted from untreated and flg22 and nlp20 treated protoplasts. **B**. Mean nuclear and cellular protein enrichment scores (Npfn and Cpfn values) of 157 proteins identified in both the nuclear and cellular protein fractions without and following flg22 treatment. Error bars denote standard error. **C**. Mean nuclear and cellular protein enrichment scores (Npfn and Cpfn values) of 73 proteins identified in both nuclear and cellular protein fractions without and following nlp20 treatment. Error bars denote standard error. Three replicates were used.

Table 9. Re-localization of mitochondrial proteins to the nucleus. Identified peptides are underlined. Transit peptides are in green. Arrow denotes the non-tryptic cleavage site.

Name	Initial part of protein sequence	Transit peptide
51 kDa subunit of complex I (AT5G08530)	1 30 52 MAPVRGILGLQRAVSIWKESNRLTPALRSF↓ STQAASTSTTPQPPPPPPPPEK	1-30
NAD-dependent malic enzyme 2 (AT4G00570)	1 24 37 MMWKNIAGLS KAAAAARTHGSRRC F↓ STAIPGPCIVHK	1-24

3.1.3 Comparison of nuclear proteomes under flg22 and nlp20 challenge

To expand on the putative set of proteins imported into the nucleus following elicitation of immunity we were interested in identifying quantitative changes in protein abundance in the nuclear proteome in the three biological scenarios (control, flg22 and nlp20). To this the PSM count used as protein quantification index (PQI) in all three biological replicate experiments were taken and performed multiple sample significance testing (ANOVA), with FDR multiples testing corrected significance threshold α=0.05, followed by post hoc test. Ninety-three proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) change in their abundance between conditions (Appendix 2, Supplementary table 3). Hierarchical clustering of these proteins showed that the three biological scenarios produced specific clusters (Figure 13 A). The protein dendrogram was divided into four main clusters as follows: proteins with decreased abundance in the nucleus following either flg22 or nlp20 stimulus (cluster 1), proteins with increased abundance in the nucleus following nlp20 stimulus (cluster 2), proteins with increased abundance in the nucleus following nlp20 and flg22 stimulus (cluster 3), proteins with increased abundance in the nucleus following flg22 (cluster 4) (Figure 13 A and Table 10). The four clusters comprise of proteins showing significant change in their abundance comparing elicited immunity to control. The proteins in these four clusters should play potential physiological roles in the nucleus during Pattern-triggered immunity (PTI). The 93 statistically significant proteins were annotated as being involved in several cellular processes such as mRNA processing, nucleotide binding, rRNA binding and include some protein families such as ribosomal proteins and prohibitin proteins.

Proteins increased in abundance upon challenge with an elicitor (proteins in clusters 2, 3 and 4) were entered into the STRING protein interaction database to identify potential physical interactions between them and infer putative functions of proteins in complexes. The abundance of all of the members of the prohibitin family which are prohibitins 1, 2, 3, 4 and 6 increased significantly (significance threshold α=0.05, FDR corrected) upon treatment with both flg22 and nlp20 (Table 10) and all interacted physically with one another (Prohibitins connected by experimental evidence in the STRING interaction network). This prohibitin complex was expanded by three members of the mitochondrial bc1 complex, MPPBETA (AT3G02090) and Cytochrome C1 family proteins AT5G40810 and AT3G27240 (Figure 13 B). The results indicate the possible function of these proteins together in the nucleus in PTI.

Nineteen ribosomal proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) change in their abundance in between untreated and flg22 and nlp20 treated samples. Ten of these increased significantly in their abundance whereas the abundance of nine decreased significantly (Table 10). The abundance of three ribosomal proteins (S4, S5, L27e) was elevated upon either flg22 or nlp20 treatment. Conversely, the abundance of five ribosomal proteins (L13, L14p, L17, L24e, and L29) increased specifically upon exposure of the protoplasts to nlp20 and the abundance of two ribosomal proteins (L16p and S11) specifically upon exposure to flg22. Protein interaction analysis showed a core set of ribosomal proteins interacting in both PTI scenarios (Figure 13 C top left panel). This core cluster however differentially expanded as proteins responding only to flg22 or nlp20 were added to the common input set (Figure 13 C bottom left and right panels). This was particularly pronounced following elicitation of PTI by flg22 (Figure 13 C right panel). It has been reported by us (Bassal et al., 2020) that ribosome composition is promiscuous dependent on cellular state and our results imply the same in the context of ribosome assembly in the nucleus.
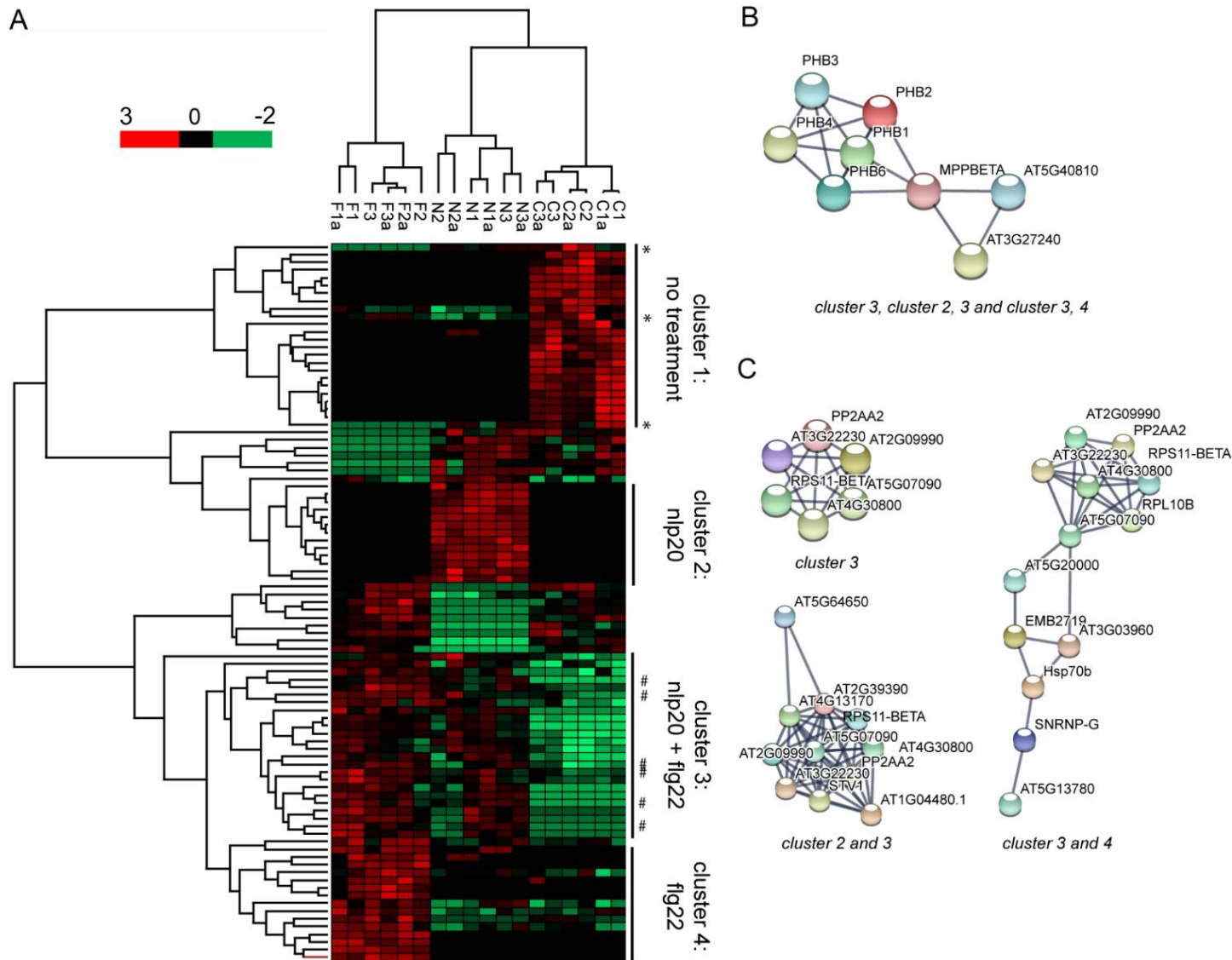
Figure 13. Proteins showing significantly changed abundance in the nuclear protein fraction following flg22 or nlp20 treatment. Multiple samples significance testing (ANOVA) was done, with FDR multiples testing corrected significance threshold α=0.05. **A**. Hierarchical cluster analysis (HCL) shows clustering of samples according to sample type. Rows represent proteins. PQI values were z-score transformed. Proteins are colored according to their abundance. * denotes decrease in abundance under effect of only flg22 or nlp20. # denotes increase in abundance under effect of flg22 only. F: flg22 treated, N: nlp20 treated, C: control (untreated), a: technical replicate. **B.** STRING database binary protein interaction network of Prohibitins with Cytochrome C generated when indicated clusters were used as input sets. **C**. STRING database interaction networks of ribosomal and associated proteins when indicated

clusters were used as input sets. String search was done using experiments, databases, co-occurrence and co-expression as interaction sources and showing only interactions between proteins in the input set. Edges in the string interaction network indicate both functional and physical protein associations.

Table 10. Four main clusters of proteins showing significant change in their abundance

List of proteins of decreased abundance in the nucleus under effect of both flg22 and nlp20 compared to control	
Cluster 1	
No treatment (control)	
AT1G54270*	eif4a-2
AT3G11500	Small nuclear ribonucleoprotein family protein
AT3G22310	putative mitochondrial RNA helicase 1
AT2G46610	RNA-binding (RRM/RBD/RNP motifs) family protein
AT2G28760	UDP-XYL synthase 6 ; UDP-XYL synthase 6 ; UDP-XYL synthase 6
AT1G07930	GTP binding Elongation factor Tu family protein
AT1G09100	26S proteasome AAA-ATPase subunit RPT5B
AT5G09500	Ribosomal protein S19 family protein
AT3G06610	DNA-binding enhancer protein-related
AT5G21160*	LA RNA-binding protein
AT2G29200	pumilio 1
AT2G31410	unknown protein
AT2G36410	Family of unknown function (DUF662)
AT4G11840	phospholipase D gamma 3
AT5G43070	WPP domain protein 1
AT3G28900	Ribosomal protein L34e superfamily protein
AT2G40010	Ribosomal protein L10 family protein
AT5G18380	Ribosomal protein S5 domain 2-like superfamily protein
AT5G15520	Ribosomal protein S19e family protein
AT5G58420	Ribosomal protein S4 (RPS4A) family protein
AT2G40510	Ribosomal protein S26e family protein
AT2G40590	Ribosomal protein S26e family protein
AT5G59240	Ribosomal protein S8e family protein
AT5G54600*	Translation protein SH3-like family protein
List of proteins of increased abundance in the nucleus under effect of flg22 and nlp20 compared to control	
Cluster 4	
Flg22	
AT4G39280	phenylalanyl-tRNA synthetase
AT5G42540	exoribonuclease 2

AT1G65540	LETM1-like protein
AT2G44525	Protein of unknown function (DUF498/DUF598)
AT1G20200	PAM domain (PCI/PINT associated module) protein
AT3G48670	XH/XS domain-containing protein
AT4G27270	Quinone reductase family protein
AT5G13780	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT3G03960	TCP-1/cpn60 chaperonin family protein
AT5G48370	Thioesterase/thiol ester dehydrase-isomerase superfamily protein
AT3G02320	N2,N2-dimethylguanosine tRNA methyltransferase
AT3G02760	Class II aaRS and biotin synthetases superfamily protein
AT5G20000	AAA-type ATPase family protein
AT1G26910	Ribosomal protein L16p/L10e family protein
AT5G03520	RAB GTPase homolog 8C
AT3G20670	histone H2A 13
Cluster 3	
Flg22 & Nlp20	
AT3G15730	phospholipase D alpha 1
AT3G02090	Insulinase (Peptidase family M16) protein
AT1G16030	heat shock protein 70B
AT4G08520#	SNARE-like superfamily protein
AT3G48820	Glycosyltransferase family 29 (sialyltransferase) family protein
AT3G25800#	protein phosphatase 2A subunit A2
AT2G23930	probable small nuclear ribonucleoprotein G
AT1G03860	prohibitin 2
AT2G20530	prohibitin 6
AT4G28510	prohibitin 1
AT5G40770	prohibitin 3
AT3G27280	prohibitin 4
AT5G12290	dgd1 suppressor 1
AT3G06860	multifunctional protein 2
AT3G27240#	Cytochrome C1 family
AT5G40810#	Cytochrome C1 family
AT3G22230	Ribosomal L27e protein family
AT5G07090	Ribosomal protein S4 (RPS4A) family protein
AT2G21410	vacuolar proton ATPase A2
AT5G23740#	ribosomal protein S11-beta
AT4G30800	Nucleic acid-binding, OB-fold-like protein
AT2G09990	Ribosomal protein S5 domain 2-like superfamily protein
AT5G14680#	Adenine nucleotide alpha hydrolases-like superfamily protein
Cluster 2	
Nlp20	

AT3G53650	Histone superfamily protein
AT1G17730	vacuolar protein sorting 46.1
AT1G04480	Ribosomal protein L14p/L23e family protein
AT2G39390	Ribosomal L29 family protein
AT4G13170	Ribosomal protein L13 family protein
AT1G20260	ATPase, V1 complex, subunit B protein
AT1G09630	RAB GTPase 11C
AT4G18430	RAB GTPase homolog A1E
AT4G27680	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT4G38780	Pre-mRNA-processing-splicing factor
AT3G53020	Ribosomal protein L24e family protein
AT5G64650	Ribosomal protein L17 family protein
AT4G02520	glutathione S-transferase PHI 2
AT4G16120	COBRA-like protein-7 precursor
* decrease in abundance under effect of only flg22 or nlp20	
# increase in abundance under effect of flg22 only	

3.2 Investigation of the nuclear proteome of *Arabidopsis thaliana* rosette leaves using proximity-dependent labelling (TurboID)

In this approach I set out to produce a high-quality draft catalog of the *A. thaliana* nuclear proteome based on proximity dependent labeling coupled with MS. Beyond this the first steps were taken towards identifying putative candidate proteins imported into the nucleus under flg22 treatment. Additionally, the quantitative changes in protein abundance in the nuclear proteome, under PTI elicited and un-elicited conditions, were investigated.

The first approach, used in this project, was based on isolation of nuclei from protoplasts of *A. thaliana* cell culture followed by devising an enrichment score to assess the purity of the preparation and curate the nuclear proteome. Here an alternative approach was used to circumvent the inevitable contaminations by using the TurboID to label the nuclear proteins in the nucleus without the need for nuclei isolation. In contrast to the cell culture experiment, the nuclear proteome of adult *A. thaliana* rosette leaves was investigated. In addition, in the frame of nuclear import under PTI, cycloheximide was used to differentiate between the *de novo* (newly) synthesized proteins and proteins that were already present and were more likely trafficked to the nucleus.

TurboID was tested and validated as a successful for proximity labeling of proteins in plants (Mair et al., 2019). They were able to generate stable *A. thaliana* lines expressing the TurboID

construct under a UBQ10 promoter. The expressed construct constituted mainly of TurboID, YFP and NLS (UBQ10pro: TurboID-YFP-NLS) (Appendix 2, Supplementary figure 2). The TurboID-YFP-NLS was expressed and targeted to the nucleus, which allowed labeling of nuclear proteins.

Seeds from the *A. thaliana* transgenic line (UBQ10pro: TurboID-YFP-NLS) expressing the essential components of the labeling systems were propagated and will be referred to as (T). Even though the T-line was positively validated, it was further checked upon arrival for the expression of the TurboID construct by recording the YFP signal by western blotting and by microscopy. Plants were grown on soil and young leaves were examined from the lower part of the leaves using confocal microscopy to check for localization of expressed TurboID in the nucleus. As shown in Figure 15, a strong YFP signal was detected in the nucleus indicating localization of TurboID. Ground frozen leaves were resuspended in sample buffer to be analyzed by western blot. A strong signal was detected at the expected size of about 65 kDa. This indicates successful expression of TurboID. However, it was observed that some plants showed no expression as in T5 and T6 (Figure 14). This suggests the silencing of the expression in some plants. This was also observed by (Mair et al., 2019). Therefore, for all of the following experiments each plant has to be checked individually for the expression of construct by checking YFP signal with western blot and microscopy.
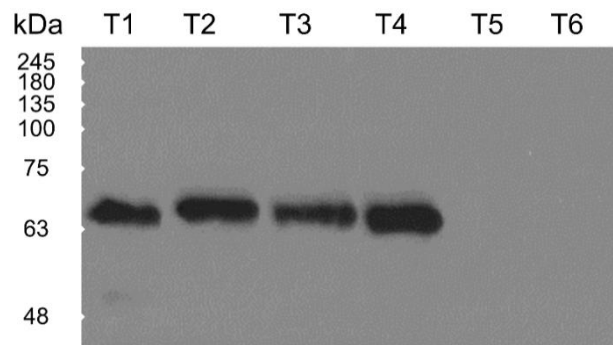
Figure 14. Checking the transgenic line T for TurboID construct expression by western blot. Western blot analysis was done by resuspending ground frozen leaf material with SDS sample buffer. The TurboID construct protein has a predicted size of about 65 kDa. Immunodetection was done using Anti-GFP polyclonal antibody and ECL prime amersham western blotting detection reagent. T: transgenic plant expressing (UBQ10pro: TurboID-YFP-NLS), T1-T6: different plants.
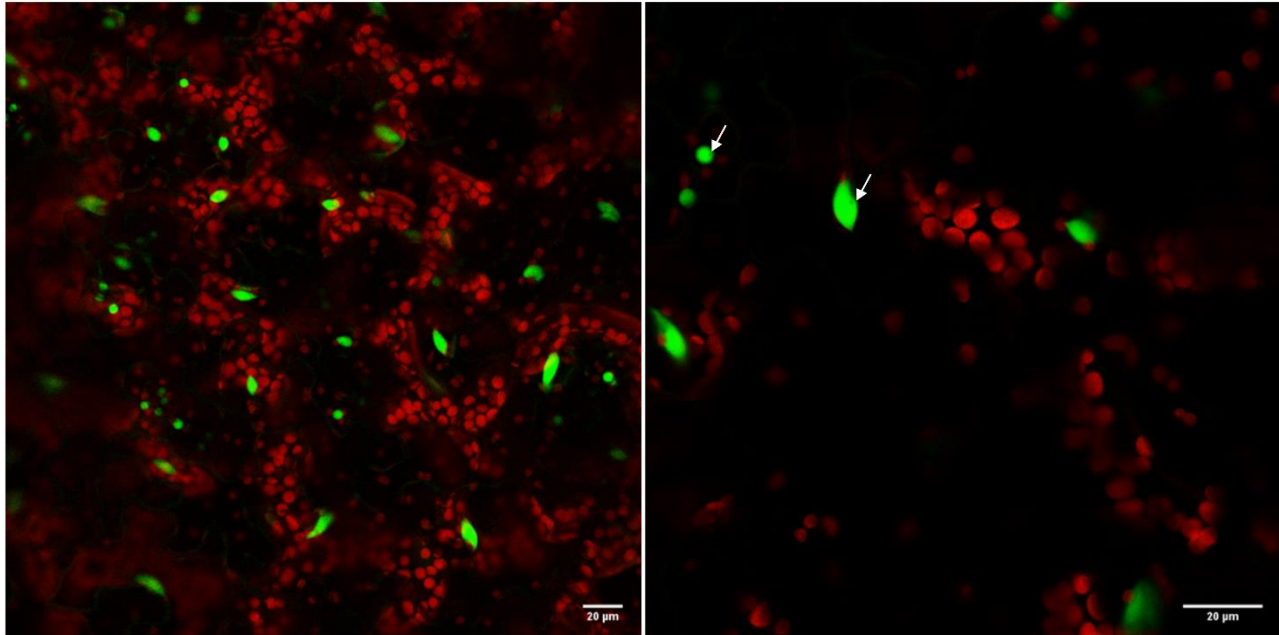
Figure 15. Checking the transgenic line T for TurboID construct expression and nuclear localization by confocal microscopy. The lower side of young leaves were examined by confocal laser scanning microscopy LSM 900 with 410-545 nm for YFP detection and 635 – 640 nm for chlorophyll detection. Red: chloroplast, yellowish green: nucleus. Scale bar: 20 µm. White arrow points to nucleolus. Several plants were checked and these are representative figures.

3.2.1 Defining the *Arabidopsis thaliana* nuclear proteome.

The method, used in this section, was developed and optimized based on protocols used in (Branon et al., 2018, Mair et al., 2019). The method is divided into several steps: plant screening for YFP signal, biotin treatment, extraction of total proteins, enrichment of biotinylated proteins and MS/MS analysis. The main steps are summarized in the following scheme (Figure 16).
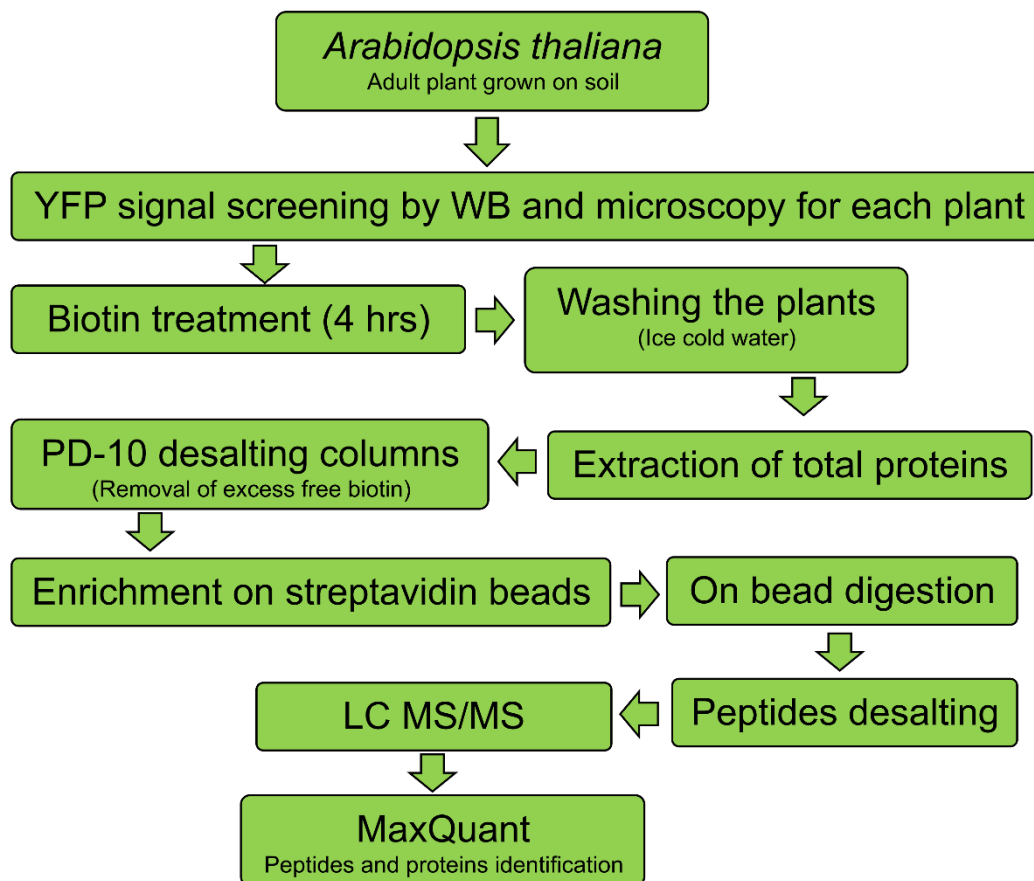
Figure 16. Summarized workflow of proximity-dependent labeling. *A. thaliana* plants were grown on soil under short-day conditions. Plants were screened for YFP fluorescence and then treated with biotin for 4 h, followed by a washing step. Total proteins were extracted and then excess free biotin was removed by gel filtration. Biotinylated proteins were enriched on streptavidin beads followed by on bead digestion and finally the dissolved desalted peptides were analyzed with liquid chromatography mass spectrometry (LC-MS) using a Data-Dependent Acquisition (DDA) scan strategy and identified and quantified by MaxQuant.

In order to investigate the nuclear proteome in *A. thaliana* adult rosette leaves, transgenic plants (T) expressing UBQ10pro: TurboID-YFP-NLS and Col-0 plants (WT), were grown on soil and after few weeks all transgenic plants were screened for expression and localization of the TurboID construct by western blot and fluorescence microscopy. Only the plants that have an YFP signal on both western blot and in microscopy were retained. Three of these plants were randomly selected and used for subsequent procedures. As shown in Figure 17 B, the three selected plants showed strong western blot signal at the expected molecular weight, whereas the WT did not show any signal. Also, these plants showed YFP signal when analyzed by

fluorescence microscopy (Figure 17 A). At the age of around nine weeks, the selected (T) and WT plants were treated with 50 μm biotin for 4 h followed by a washing step with cold water to stop biotin activation and remove excess free biotin. Total protein extraction was performed on the frozen ground plant rosette leaves. Protein extracts were mixed with SDS sample buffer and western blot analysis using Streptavidin-Pod conjugate antibody was done to check for success of the biotin treatment. As shown in Figure 17 C, the transgenic plants (T) showed a very strong signal throughout the range of molecular weight when compared to the WT. This indicates the successful and abundant biotinylation of proteins. WT showed few signals, which denote naturally biotinylated proteins.
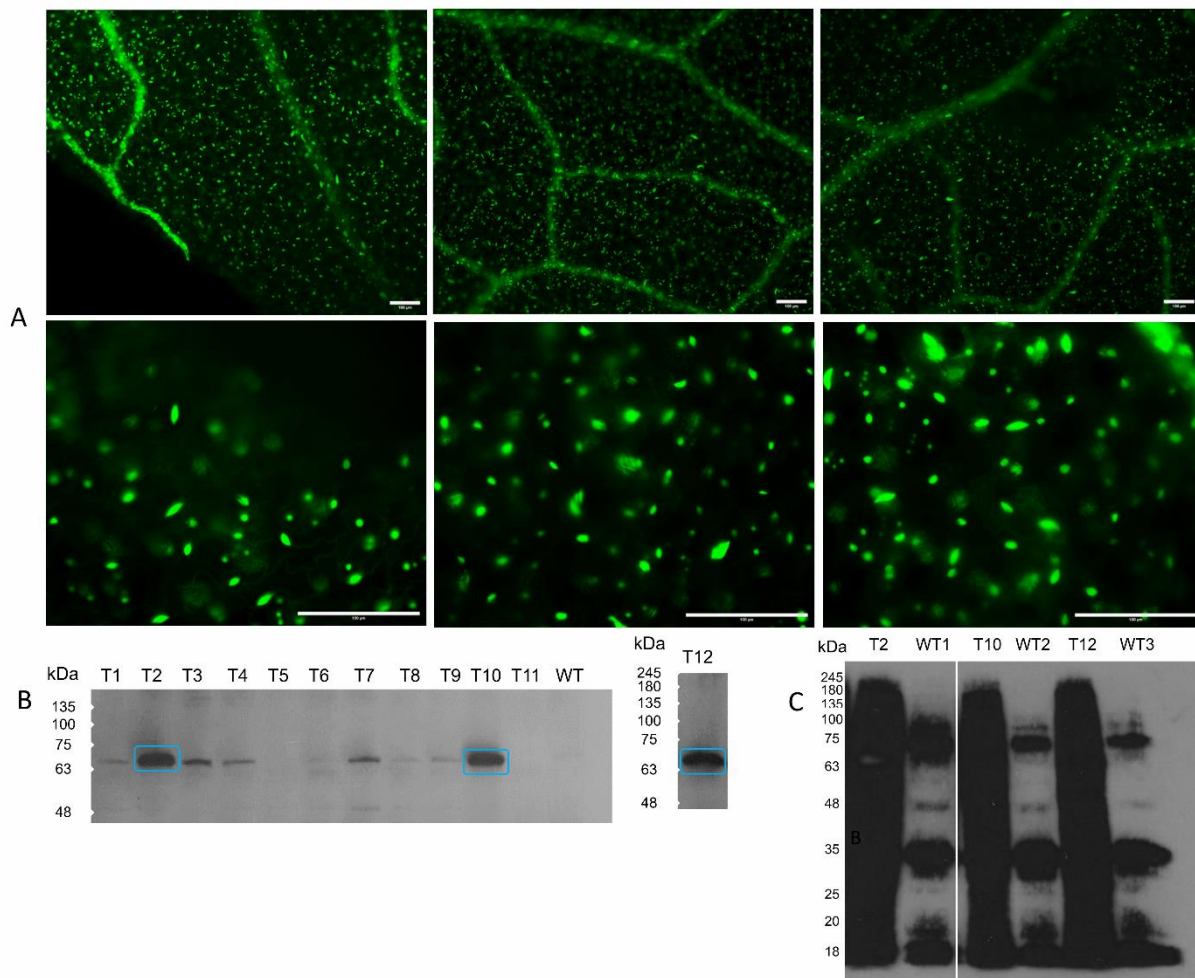
Figure 17. (A and B) Screening the transgenic plants (T) for the TurboID (by YFP fluorescence and western blot). (C) checking the success of biotin treatment. A. The abaxial sides of three weeks old leaves were examined for YFP-fluorescence with an Axioplan2

imaging fluorescence microscope employing an YFP filter. Green = nucleus; Scale bar = 100 µm. Representative images from three plants are shown. **B**. Immunodetection of YFP in leaf material from T-lines. The analysis was done by resuspending ground frozen leaf material in SDS sample buffer, followed by immunodetection using an anti-GFP polyclonal antibody and ECL prime amersham western blotting detection reagent. The TurboID construct protein has a predicted size of about 65 kDa. T: transgenic plant expressing (UBQ10pro: TurboID-YFP-NLS), T1-T12: different plants, WT: Col-0 wild type. Signals marked with blue rectangles were the selected plants for further analysis. **C**. Analysis was done by mixing protein extracts with SDS sample buffer. Immunodetection was done using Streptavidin-Pod conjugate antibody and Pierce ECL plus western blotting substrate. Volume containing about 15 µg of total proteins were applied to each lane. T = transgenic plant, WT = Col-0 wild type.

Before enrichment of biotinylated proteins, the protein extracts were applied to PD-10 gel filtration columns to ensure removal of excess free biotin that could interfere with binding. Then the protein extracts were incubated with Dynabeads Myone Streptavidin C1 for 16 h. This was followed by washing and on bead digestion with trypsin. Then the desalted dissolved peptides were analyzed with liquid chromatography mass spectrometry (LC-MS) using a Data-Dependent Acquisition (DDA) scan strategy. The MS analysis of Transgenic samples (T) identified 2105 protein groups in total (Appendix 1, Supplementary file 2- Table 1). This set of proteins was curated to eliminate non-biotinylated proteins that bound unspecifically to the beads and also naturally biotinylated proteins. Therefore, three biological replicates of WT plants were used as controls, and the identified set of proteins was further filtered to eliminate proteins identified in any of the three WT replicates (even if identified in only one replicate). Based on this analysis, 81 proteins was identified in both WT and T and were removed leading to a curated nuclear proteome set of 2024 protein groups (Figure 18) and (Appendix 1, Supplementary file 2-Table 3).
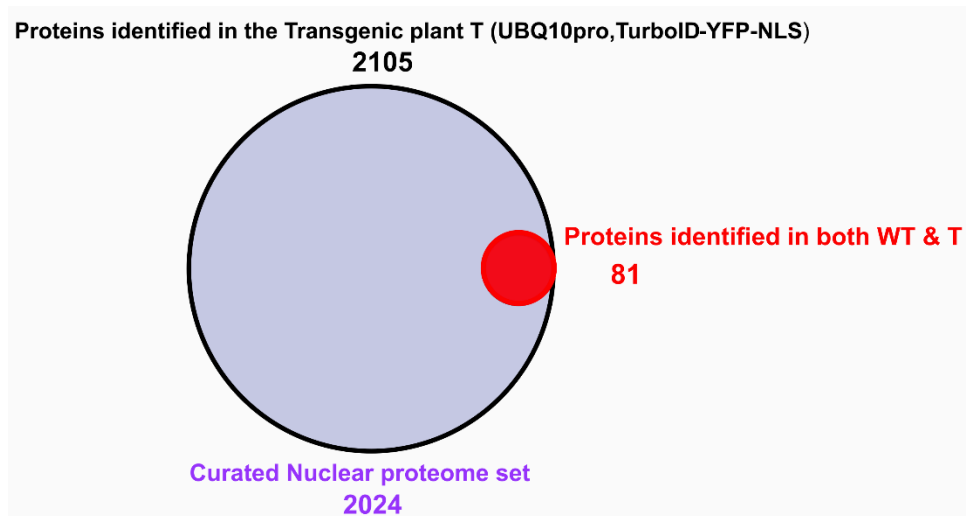
Figure 18. Curation of nuclear proteome. The identified 2105 protein groups set was filtered to remove the common proteins in WT representing the proteins that were bound to beads unspecifically and as well as the naturally biotinylated proteins leaving a set of 2024 curated nuclear protein groups. Three biological replicates were used.

To further validate the nuclear proteomes, the curated protein lists were analyzed with the DAVID Bioinformatics resources 6.8 gene ontology tool (Huang et al., 2009a, Huang et al., 2009b) and LOCALIZER (Sperschneider et al., 2017), a software that predicts organelle subcellular localization by searching for targeting sequences such as NLS in protein primary structure. Also our experimentally determined proteomes were compared with previously published nuclear / sub-nuclear proteomes and nuclear localized proteins by FP (Bae et al., 2003, Pendle et al., 2005, Calikowski et al., 2003, Bigeard et al., 2014, Sakamoto and Takagi, 2013, Chaki et al., 2015, Goto et al., 2019, Palm et al., 2016, Mair et al., 2019, Tang et al., 2020, Hooper et al., 2017). As a result, 93 % of the nuclear proteins consisted of either experimentally verified nuclear proteins or proteins annotated / predicted to be localized in the nucleus (redundancy was removed), (Figure 19 A). This underscores the high quality of the nuclear proteome.

The DAVID bioinformatics tool was used to further annotate the nuclear proteins and classify them according to their function. 1325 proteins were classified initially as belonging to the nucleus. This set was then further input into DAVID. These proteins were categorized into 143 clusters and 9 main protein classes (Figure 19 B and appendix 1, Supplementary file 2- Table 4). The nine main classes were: transcription (322 proteins), metal binding (139 proteins),

nucleotide binding (87 proteins), RNA binding (60 proteins), Transport including proteins (63 proteins), mRNA processing (27 proteins), translation initiation and elongation activity (22 proteins), cell division (20 proteins) and DNA repair (11 proteins).
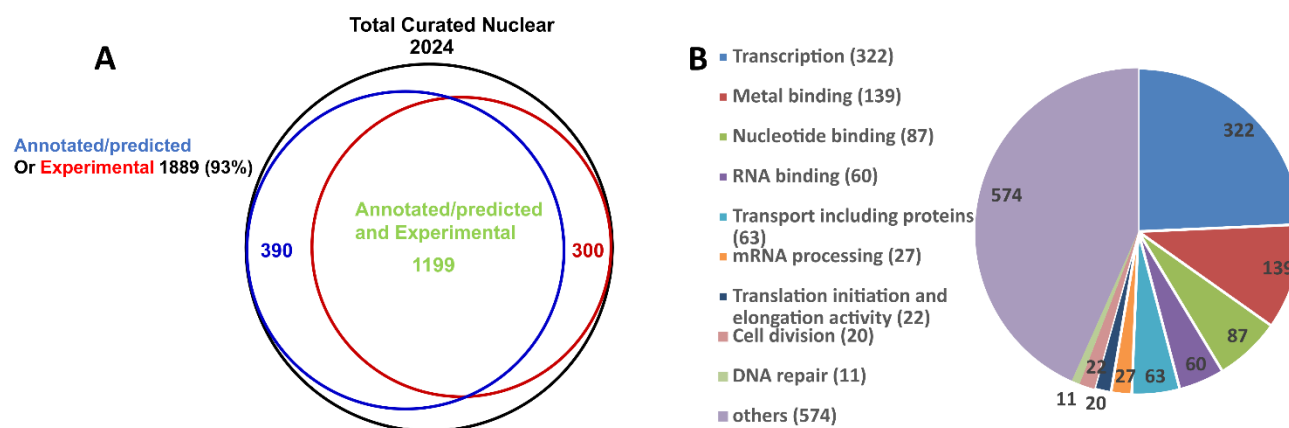
Figure 19. Defining the nuclear proteome (proximity-dependent labeling). A. VENN diagram showing sets of experimentally validated or annotated/predicted nuclear proteins, comprising our data set. **B**. DAVID gene ontology classification of 1325 proteins annotated as nuclear proteins by DAVID bioinformatics tool. Three biological replicates were used.

Proteins annotated as related to the process of transcription, were classified into families, as shown in (Table 11) and Appendix 2, Supplementary table 4 . The top three transcription factors families pertaining to the number of proteins identified were WRKY (19 proteins), bHLH (19 proteins) and bZIP (16 proteins). Nuclear envelope proteins were also identified. In control conditions, ten proteins were reported to be localized in the nuclear envelope, inner and outer membrane based on FP experimental evidence in SUBA4. Moreover, 130 proteins were reported to be found at the same locations based on FP and MS experimental evidence in SUBA4. All together our results constitute a high-quality catalog of the nuclear proteome of *A. thaliana*.

Table 11. Proteins annotated to transcription process were classified into families.
UP_KEYWORDS category was used in the David bioinformatics resources annotation

Families	Number of proteins	Families	Number of proteins
(Transcription factors)		**(Transcriptional factors)**	
bHLH	19	ZF-HD	3
WRKY	19	ARF	2
bZIP	16	ARR-B	2
Trihelix	11	B3	2
C2H2	10	CAMTA	2
G2-like	10	E2F/DP	2
HSF	8	ERF	2
TCP	8	GATA	2
Dof	7	HB-other	2
MYB_related	7	NF-YC	2
NAC	7	SBP	2
MYB	5	AP2	1
TALE	5	BES1	1
BBR-BPC	4	C3H	1
GeBP	4	HB-PHD	1
GRAS	4	Nin-like	1
HD-ZIP	3	VOZ	1
		WOX	1

Due to the complexity of the experimental set up in this approach, I wanted to preliminary check the repeatability of the method. Therefore, the $\log_2$ transformed LFQ values (protein quantification index, PQI) of proteins quantified in all three replicates were looked up and the relative standard deviation (% RSD) were calculated for these proteins. It was found that 88 % of these proteins have % RSD < 5 and highest % RSD found was 15. This indicates the low variability of LFQ values of these proteins in between the three replicates and shows that the method has an acceptable reproducibility and is suitable for quantification.

3.2.2 Rearrangement of the nuclear proteome under PTI.

Flg22 was used as stimulus to trigger PTI in *A. thaliana*. The rosette leaves were infiltrated with flg22 solution and left for 1 h and then the rosette leaves where subjected to biotin treatment for 4 h and was frozen (in total 5 h after flg22 treatment). Procedures were then performed as mentioned in the scheme (Figure 16) and under 2.2.2 in the methods section. The infiltration method is a standard procedure for application of flg22. Even though, I wanted to check for the success of flg22 treatment because of the complexity of experimental set up and the possible

interference with biotin treatment. Therefore, trial plants were used and were treated exactly the same as in the planned procedures until the end of the biotin treatment step. These plants were used in a small qPCR experiment to test PAL1 (Phenylalanine ammonia lyase1 and TSA1 (tryptophan synthase alpha chain). These two marker genes are known to be induced after PTI and were previously investigated in our lab (Bassal et al., 2020). A relative quantification method was used as explained in (Schmittgen and Livak, 2008) using UBC21 and PP2A as internal standards. As shown in (Figure 20) both PAL1 and TSA 1 were significantly (Two tailed t-test was used at α of 0.05) induced after flg22 treatment when compared to the water treated control plants indicating that flg22 infiltration was able to trigger the PTI.
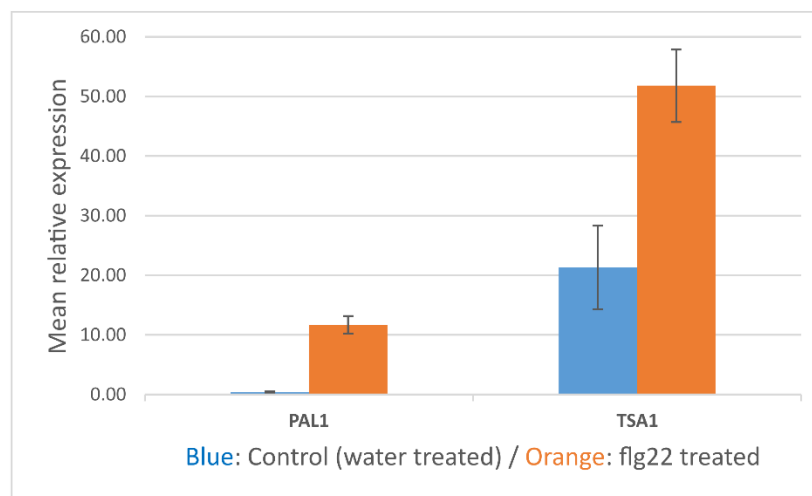
Figure 20. qPCR experiment for PAL1 and TSA1 after flg22 treatment. *A. thaliana* rosette leaves were infiltrated with flg22 and with water as control and the mean relative expression of PAL1 and TSA1 genes was measured. UBC21 and PP2A were used as housekeeping genes and as internal standards. Two tailed t-test was used at α of 0.05. Error bars denote standard error. Three biological replicates were used.

A large scale experiment was planned to identify LC-MS-based candidate proteins for import to the nucleus from other organelles and to investigate the quantitative changes of protein abundance in the nucleus after flg22 treatment. Therefore, transgenic plants (T) expressing UBQ10pro: TurboID-YFP-NLS and Col-0 plants (WT) were grown on soil. Screening was done the same as above and only plants that had a positive YFP signal on western blot and in microscopy were retained. Twelve transgenic plants were used in this experiment and their western blot results and representative microscopy images are shown in (Figure 21 A and B). These plants were randomly assigned to four groups. The first one was the water treated control (W), the second one flg22 treated (F), the third flg22 and cycloheximide treated (FC) and the

last one cycloheximide treated (C). The rosette leaves were infiltrated with either flg22 or/and cycloheximide and were left for 1 h and then the rosette leaves where subjected to biotin treatment for 4 h and then was frozen (in total left for 5 h after flg22 treatment). In the case of cycloheximide treatment, the cycloheximide was additionally supplemented with biotin solution to ensure continuous exposure during the entire time period. Protein extracts were analyzed with western blot as above and showed the success of the biotin activation step (see Figure 21 C for representative western blots).
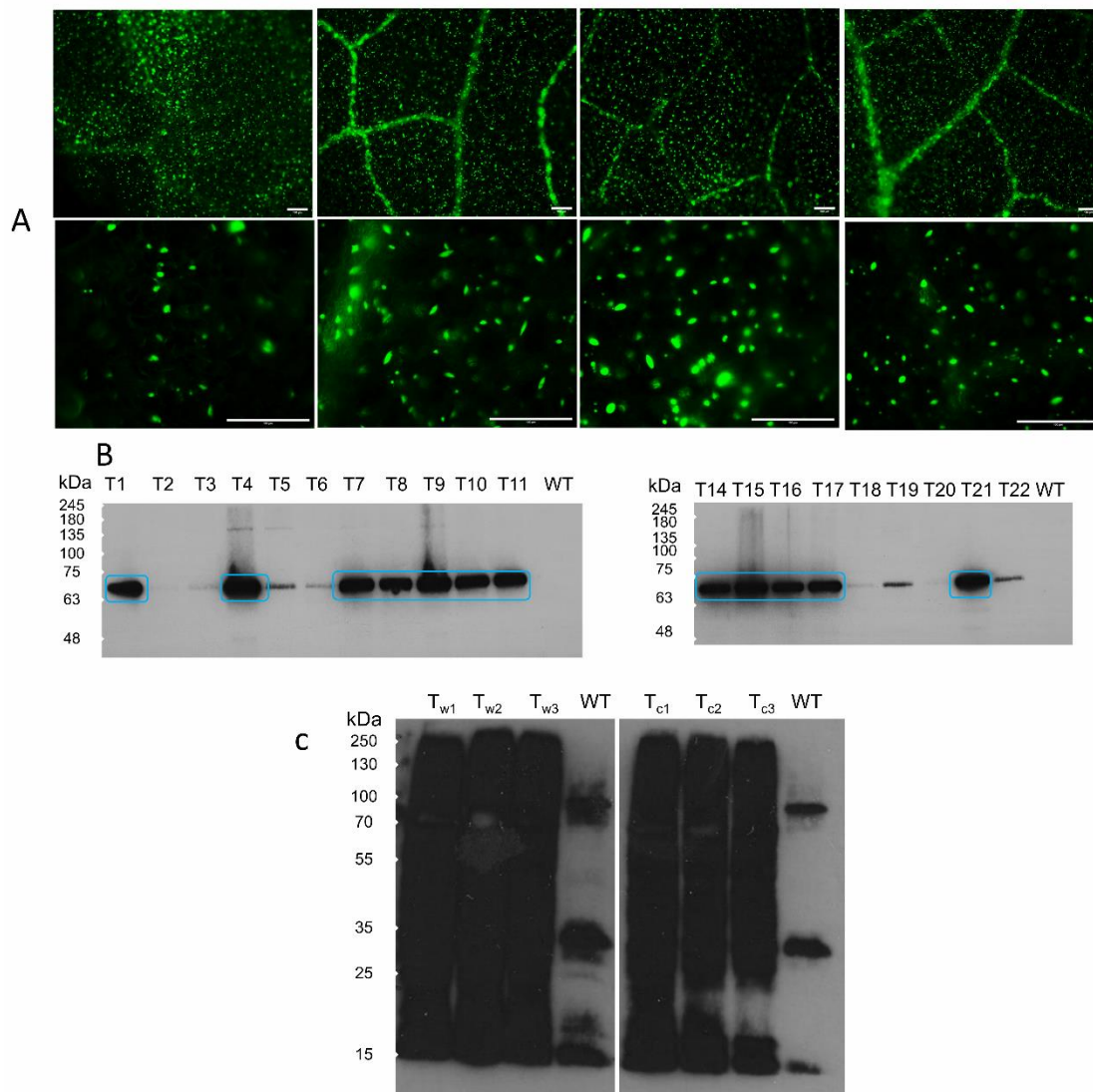
Figure 21. (A and B) Screening the transgenic plants (T) for the TurboID (by YFP fluorescence and western blot). (C) Checking the success of biotin treatment step. A. The abaxial sides of three weeks old leaves were examined for YFP-fluorescence with an Axioplan2 imaging fluorescence microscope employing an YFP filter. Green = nucleus; Scale bar = 100

μm. Representative images from the entire set of plants are shown. **B**. Immunodetection of YFP in leaf material from T-lines. The analysis was done by resuspending ground frozen leaf material in SDS sample buffer, followed by immunodetection using an anti-GFP polyclonal antibody and ECL prime amersham western blotting detection reagent. The TurboID construct protein has a predicted size of about 65 kDa. T: transgenic plant expressing (UBQ10pro: TurboID-YFP-NLS), T1-T22: different plants, WT: Col-0 wild type. Signals marked with blue rectangle were the selected plants. **C**. Analysis was done by mixing protein extracts with SDS sample buffer. Immunodetection was done using Streptavidin-Pod conjugate antibody and Pierce ECL plus western blotting substrate. Volume containing about 15 μg of total proteins were applied to each lane. T = transgenic plant, WT = Col-0 wild type, w = water treated, c = cycloheximide treated. Representative western blots are shown.

Aliquots of the protein extracts were kept a side to be used later in a confirmatory experiment to check for the effectiveness of cycloheximide treatment (see under page 79). The rest of protein extracts were used the same as mentioned before (Figure 16) the biotinylated proteins were enriched and the final desalted peptides were analyzed with liquid chromatography mass spectrometry (LC-MS) using a Data-Dependent Acquisition (DDA) scan strategy. The MS analysis of Transgenic samples (T) identified 2035 protein groups in total. This set of proteins were curated using the same approach as mentioned under 3.2.1 using a list of, proteins identified in WT leading to a curated set of 1940 proteins (Appendix 1, Supplementary file 2-Tables 5-7).

In order to identify candidate proteins imported to the nucleus upon flg22 challenge, the curated list of nuclear proteins was compared between all conditions (W, F, FC and C). Among these proteins, 51 proteins were absent (not identified in all three replicates) in the nucleus under water treated control condition (W) and were present (identified in at least two of three replicates) in the nucleus under flg22 treated condition (F) (Figure 22) and (Appendix 2, Supplementary table 6). This means that in these experiments considering our detection limits, these proteins appeared in the nucleus after elicitation of immunity. These 51 proteins could either originate from the cytosol or other organelles and be imported to the nucleus upon PAMP perception, or they could be newly synthesized and then imported. In order to investigate this further, cycloheximide was used to inhibit protein synthesis allowing to differentiate between newly synthesized proteins and proteins trafficked to the nucleus from the cytosol or other organelles. In consequence, 21 proteins were found to be absent from the nucleus under water treated control condition (W), but were present in both flg22 treated condition (F) and flg22 + cycloheximide treated condition (FC) (Figure 22) and (Appendix 2, Supplementary table 7). Assuming the cycloheximide treatment was successful, these candidate proteins might have

been trafficked to the nucleus from the cytosol or some other organelle upon elicitation of immunity, possibly indicating a physiological role in plant defence. Bioinformatics analysis was done in order to investigate the possible subcellular localization (other than the nucleus) of these proteins. SUBA4 was used to check for localization inferred by experimental evidence and Localizer and subacon were used for prediction. In addition, the GO cellular component (annotation) from TAIR was also used (Table 12). Moreover, 90 proteins were absent in nucleus under water treated control condition (W) and were present in nucleus under cycloheximide treated condition (C) (Appendix 2, Supplementary table 8). This indicates that these proteins were imported to the nucleus and could be trafficked from cytosol or other organelles after the cycloheximide exerted stress.

Absent in nucleus (W) and present in nucleus (F)

51

Absent in nucleus (W), present in nucleus (F) and present in nucleus (FC)

21
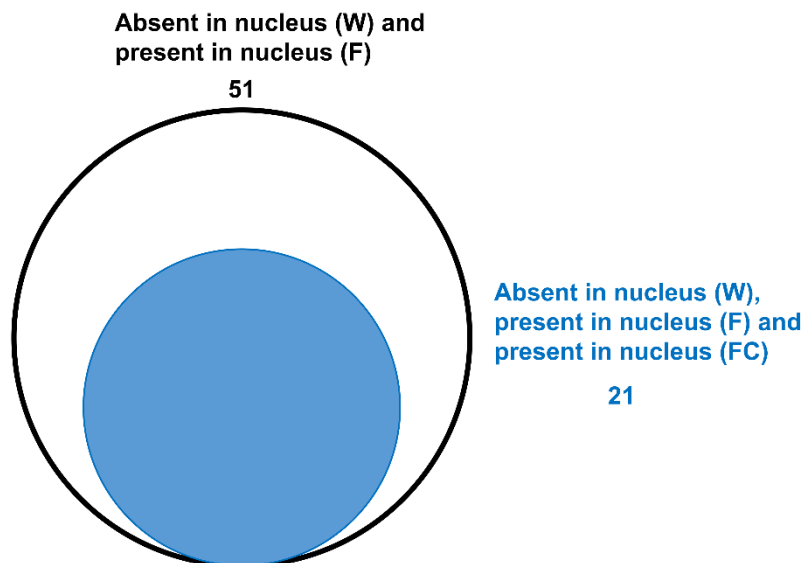
Figure 22. Protein import into the nucleus after flg22 treatment. 51 candidate proteins, for import into the nucleus under flg22 treatment, were filtered to produce a set of 21 candidate proteins which could be trafficked from cytosol or other organelles.

Table 12. Candidate proteins for import to the nucleus after possible trafficking from other organelles in PTI. Only the known other subcellular locations than the nucleus are shown.

| | Experimental | | Prediction/annotation | | |
| | SUBA4 | | | | |
	FP	MS/MS	Subacon	Localizer	GO annotation on TAIR
AT1G09230			Cytosol		
AT3G13470		Golgi Mitochondrion P. membrane plastid Vacuole	Plastid	Chloroplast	Chloroplast Chlo. Envelope Chlo. stroma Cyt. ribosome Mitochondria
AT1G71697			Cytosol		Cytoplasm
AT1G79690		Cytosol Extracellular Vacuole	Cytosol		Cytoplasm Cytosol Mitochondria Vacuole
AT2G45880			Cytosol		Extracellular
AT3G07720	Cytosol	Cytosol Extracellular	Cytosol		Cytosol Cytoplasm
AT3G08530		Cytosol extracellular Golgi P. membrane Plastid Vacuole	Plasma membrane		Golgi Chloroplast P. membrane Plasmodema,
AT3G12200	Cytoskeleton	Cytoskeleton			Cytoplasm P. membrane
AT3G19080					Plasmodesma
AT4G34660	Cytosol	Cytosol Golgi P. membrane	Cytosol		Cell plate Cytoplasm Cytosol Endosome P. membrane
AT5G16300		Cytosol Golgi Vacuole	Golgi		Golgi Plastid
AT5G23060	Mitochondrion P. membrane Plastid	Golgi Mitochondrion P. membrane Plastid Vacuole	Plastid	Chloroplast Mitochondria	Chloroplast Chlo thylakoid Mitochondria Plastid Thylakoid
AT5G42190	Cytosol	Extracellular	Cytosol		Cytoplasm Cytosol Plastid
Accession	**Description**				
AT1G09230	U11/U12-65K ribonucleoprotein				
AT3G13470	TCP-1/cpn60 chaperonin family protein				
AT1G71697	ATCK1, CK, CK1	choline kinase 1				

AT1G79690	atnudt3, NUDT3 | nudix hydrolase homolog 3
AT2G45880	BMY4, BAM7 | beta-amylase 7
AT3G07720	Galactose oxidase/kelch repeat superfamily protein
AT3G08530	Clathrin, heavy chain
AT3G12200	AtNek7, Nek7 | NIMA-related kinase 7
AT3G19080	SWIB complex BAF60b domain-containing protein
AT4G34660	SH3 domain-containing protein
AT5G16300	COG1/Vps51/Vps67 family (components of vesicular transport) protein
AT5G23060	CaS | calcium sensing receptor
AT5G42190	ASK2, SKP1B | E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 family protein

In order to investigate the quantitative changes in protein abundance in the nuclear proteome In PTI, the LFQ values (protein quantification index, PQI) in all three biological replicates for all conditions were taken and used to perform multiple sample significance testing (ANOVA), with FDR multiples testing corrected significance threshold α=0.05, followed by post hoc test. Based on that, 174 proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) change in their abundance in the nucleus in between conditions (W, F, FC and C). Moreover, 94 proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) change in their abundance between cycloheximide treated condition (C) and water treated control (W) indicating the strong effect of cycloheximide on nuclear proteome rearrangements. A further 22 proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) change in their abundance between flg22 treated condition (F) and water treated control (W) (Appendix 2, Supplementary table 9). The 22 proteins were extracted and categorized into two main clusters. Proteins with decreased abundance in the nucleus following flg22 stimulus (cluster 1) and proteins with increased abundance in the nucleus following flg22 stimulus (cluster 2) (Figure 23 and Table 13). The two clusters comprise proteins showing a significant change in their abundance comparing elicited immunity to control. The proteins in these two clusters could have a potential physiological roles in the nucleus during Pattern-triggered immunity (PTI). According to their annotation, the 22 proteins with a statistically significant change in abundance are involved in different physiological processes, such as transcription, DNA binding, metal binding or responses to infections. They include proteins, such as WRKY40, WRKY72, phosphoproteins, calcium binding EF-hand family protein (CML40) and Glutathione S-transferase (GSTL3).

Table 13. Two main clusters of proteins showing significant change in their abundance between flg22 and control conditions.

List of proteins of increased abundance in the nucleus under effect of flg22 compared to control	
Cluster 2	
flg22 treated	
AT1G17210	ATILP1, ILP1 | IAP-like protein 1
AT1G69800	Cystathionine beta-synthase (CBS) protein
AT1G80840	WRKY40, ATWRKY40 | WRKY DNA-binding protein 40
AT2G35830	unknown protein
AT2G37970	SOUL-1 | SOUL heme-binding family protein
AT2G40140	CZF1, ZFAR1, SZF2, ATSZF2 | zinc finger (CCCH-type) family protein
AT2G44370	Cysteine/Histidine-rich C1 domain family protein
AT3G01830	Calcium-binding EF-hand family protein
AT5G02790	GSTL3 | Glutathione S-transferase family protein
AT5G15130	WRKY72, ATWRKY72 | WRKY DNA-binding protein 72
AT5G25260	FLOT2, SPFH/Band 7/PHB domain-containing membrane-associated protein family
AT5G53000	TAP46 | 2A phosphatase associated protein of 46 kD
AT5G62000	ARF2, ARF1-BP, HSS, ORE14 | auxin response factor 2
List of proteins of decreased abundance in the nucleus under effect of flg22 compared to control	
Cluster 1	
Water treated (control)	
AT1G75100	JAC1 | J-domain protein required for chloroplast accumulation response 1
AT2G25450	GSL-OH, GLUCOSINOLATE HYDROXYLASE
AT2G43910	ATHOL1, HOL1 | HARMLESS TO OZONE LAYER 1
AT2G45820	Remorin family protein
AT3G12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein
AT3G45190	SIT4 phosphatase-associated family protein
AT4G16330	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT4G22150	PUX3 | plant UBX domain-containing protein 3
AT4G27900	CCT motif family protein
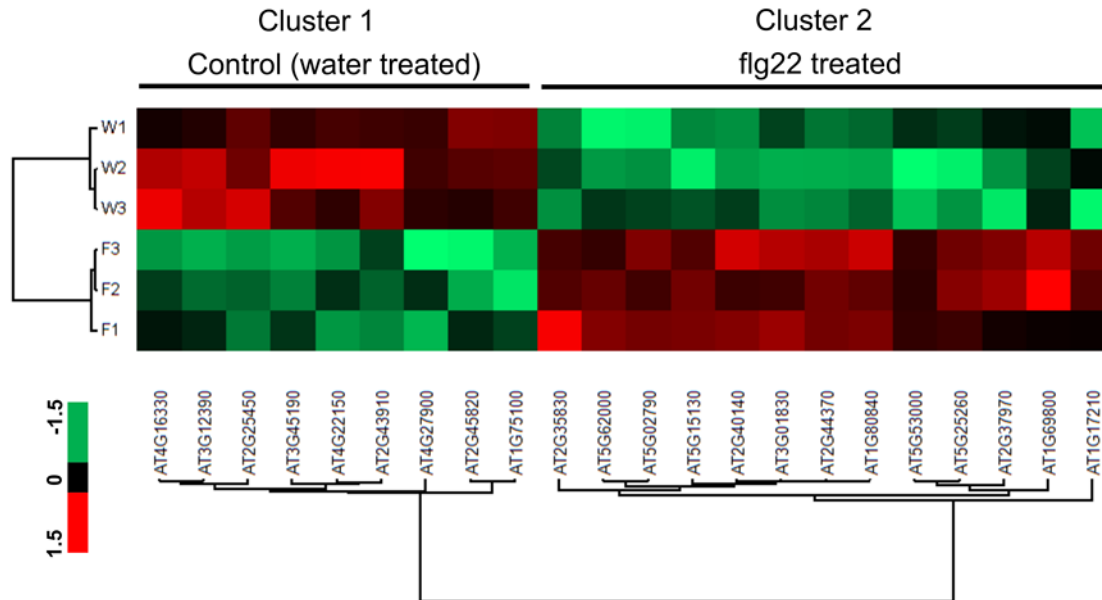
Figure 23. Proteins showing significant changes in their abundance in the nucleus following flg22 treatment compared to water treated control. Multiple sample significance testing (ANOVA) was done, with FDR multiples testing corrected significance threshold α=0.05. A. Hierarchical cluster analysis (HCL) shows clustering of samples according to sample type. PQI values were z-score transformed. Proteins are colored according to their abundance. F: flg22 treated, W: control (water treated). Three biological replicates were used.

Cycloheximide was used to inhibit protein synthesis and to allow differentiation of newly synthesized proteins and proteins trafficked to the nucleus from the cytosol or other organelles during PTI. Therefore, a confirmatory experiment was performed to assess the effectiveness of cycloheximide treatment. As mentioned on page 74, a part of the protein extracts prepared was not used for enrichment on streptavidin beads, but for analysis of total protein in the whole cell. The protein extracts were digested with FASP protocol and the desalted peptides were analyzed on MS with DDA mode. The LFQ values (protein quantification index, PQI) in all three biological replicates for the flg22 treated (F), flg22 treated + cycloheximide treated (FC) and water treated control (W) were taken and performed multiple sample significance testing (ANOVA), with FDR multiples testing corrected significance threshold α=0.05, followed by post hoc test. Based on this analysis, 61 proteins showed a statistically significant (significance threshold α=0.05, FDR corrected) increase in their abundance after the flg22 treatment when compared to water treated control (W). 93% of these proteins (57 proteins) did not show any statistically significant increase in their abundance after the flg22+cycloheximide treatment

when compared to water treated control (W) (Figure 24 and appendix 2, Supplementary table 10). Most of the 57 proteins are known to be upregulated in PTI, 25 of them were investigated by us before (Bassal et al., 2020) and 12 were involved in biotic and abiotic stress according to the TAIR database. This indicates that the cycloheximide treatment was successful and cycloheximide was able to inhibit protein synthesis after flg22 treatment in this experimental setup.
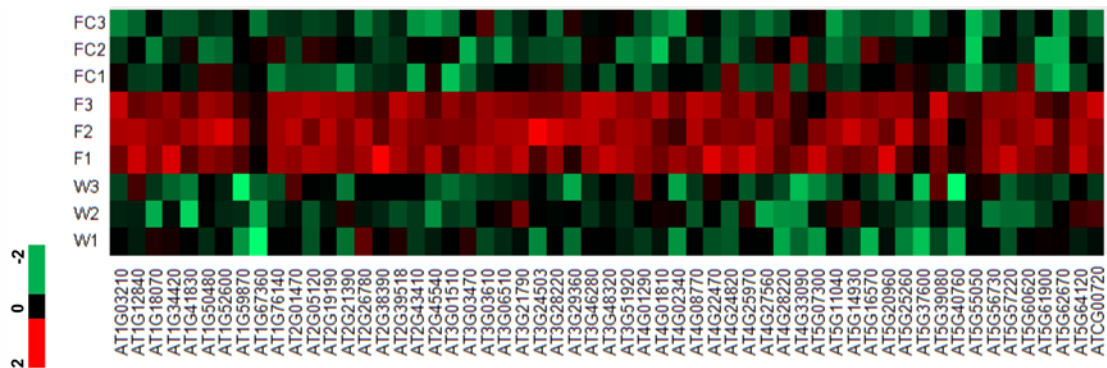
Figure 24. Confirmatory experiment to check for cycloheximide treatment effectiveness. Heat map showing the abundance of 57 proteins showing changes in abundance after exposure to flg22. Multiple samples significance testing (ANOVA) was done, with FDR multiples testing corrected significance threshold α=0.05. PQI values were z-score transformed. Proteins are colored according to their abundance. F: flg22 treated, W: control (water treated) and FC: flg22+cycloheximide treated. Three biological replicates were used.

3.3 <u>LC-MS-based candidate proteins newly identified in the nucleus and putative dual targeted proteins in *Arabidopsis thaliana*.</u>

The curated nuclear proteome list under control conditions identified with both nuclei isolation and proximity labeling approaches were compared with previously published nuclear / sub-nuclear proteomes and nuclear localized proteins by FP. This meta analysis showed (to our knowledge) that 947 proteins were newly identified (experimentally) in the nucleus by the nuclei isolation approach and 521 proteins were newly identified (experimentally) in the nucleus by proximity labeling approach. From these two protein sets of newly identified proteins, 90 proteins were identified by both approaches, and these are considered newly identified nuclear proteins with high confidence (Figure 25 and appendix 1, Supplementary file 2-Tables 8-10).
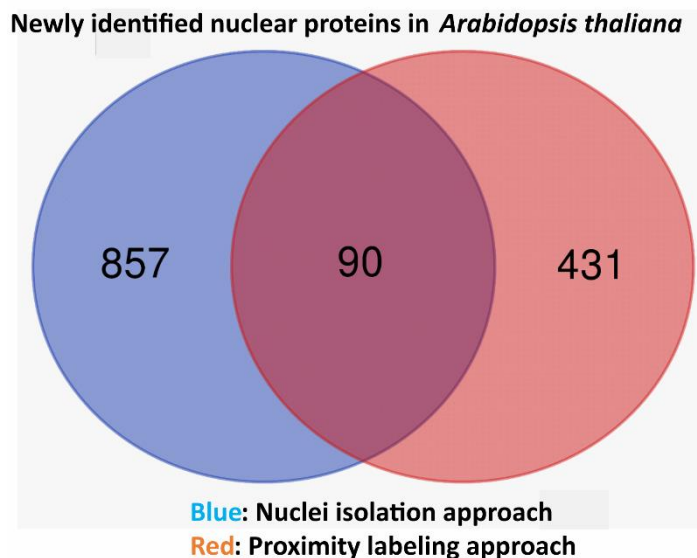
Figure 25. Newly identified (experimentally) nuclear proteins in *Arabidopsis thaliana.*

In the previous sections protein import into the nucleus and dual targeting of proteins in PTI was investigated. Additionally in this section, a bioinformatics analysis combined with our LC-MS-based analysis was done to identify proteins that potentially underlie dual targeting to the nucleus and more than one organelle under control conditions. With this type of analysis, a putative candidate list of proteins will be provided as a valuable resource for further investigations with orthogonal methods for confirmation. The curated nuclear protein list at control conditions from the nuclei isolation approach and the proximity labeling approach were combined and redundancy was removed. The combined list was analyzed with SUBA4 (Hooper et al., 2017), to check for proteins identified in previous experiments in plastids and mitochondria using FP or MS/MS. 677 proteins in the combined nuclear list were also found to be experimentally localized in the plastid and 633 in the mitochondrion, indicating their possible dual targeting to the nucleus and these two organelles. Out of the 677 putative dual- targeted plastid proteins, 316 proteins (47%) contained NLS, which were further refined to a set of 95 proteins containing both NLS and a predicted chloroplast transit peptide. In contrast, only 37% of all proteins contained in SUBA4 experimentally shown to be localized in the plastid also had an NLS, a notable lower percentile. Out of 633 putative dual targeted mitochondrial proteins, 282 proteins contained NLSs (45%) and again 91 showed both an NLS and a predicted mitochondria transit peptide (Table 14 and appendix 1, Supplementary file 2-Tables 11-12). Only 31% of all mitochondrial proteins in SUBA4 also had the NLS, again a notably lower

fraction than in our experimental set, underscoring possible dual targeting. The lists of dually targeted candidate proteins were compared with previously published nuclear and sub-nuclear proteomes (as mentioned before) and analyzed with SUBA4 to determine experimental localization in the nucleus. Among the potentially dual targeted candidate proteins, to our knowledge 69 and 83 (nucleus-plastid and nucleus-mitochondrion, respectively) proteins were not previously reported to reside in the nucleus. All of the proteins in these two sets contained NLS, according to localizer analysis. Interestingly, 37 (nucleus-plastid) and 37 (nucleus-mitochondrion) candidate proteins contained both an NLS and a predicted transit peptide (Table 14 and appendix 1, Supplementary file 2-tables 11-12), consistent with their inferred dual targeting.

Table 14. Putative dual Targeted proteins (found in both organelles)

Nucleus - Plastid		Nucleus - Mitochondrion	
Proteins with alternative sub-cellular location (SUBA4)	677	Proteins with alternative sub-cellular location (SUBA4	633
Containing NLS	316	Containing NLS	282
Containing NLS and predicted chloroplast transit peptide	95	Containing NLS and predicted mitochondrion transit peptide	91
Newly identified containing NLS	69	Newly identified containing NLS	83
Newly Identified containing NLS and predicted transit peptide	37	Newly Identified containing NLS and predicted transit peptide	37

4. Discussion

The large scale LC-MS, discovery-driven study presented in this thesis investigated the *A. thaliana* nuclear proteome after elicitation of PTI by two approaches: Nuclei isolation coupled with MS and proximity dependent labeling coupled with MS. Despite the central role of the nucleus in regulating gene expression, the plant nuclear proteome remains somewhat understudied when compared to the mammalian and human nuclear proteome (Thul et al., 2017, Go et al., 2021, Petrovská et al., 2015, Yin and Komatsu, 2016, Jez et al., 2021). In *A. thaliana* an early study using two-dimensional gel-electrophoresis (2-DE) characterized the nuclear proteome upon cold stress but the coverage was limited (Bae et al., 2003). Another six studies have reported nuclear sub-proteomes such as nucleolar proteins, nuclear matrix or chromatin associated proteins and nuclear envelope (Bigeard et al., 2014, Calikowski et al., 2003, Chaki et al., 2015, Pendle et al., 2005, Sakamoto and Takagi, 2013, Tang et al., 2020). Only recently, however three studies have described the *A. thaliana* nuclear proteome comprehensively employing LC-MS (Goto et al., 2019, Palm et al., 2016, Mair et al., 2019). This project, nuclei isolation approach which was published (Ayash et al., 2021) and the proximity labeling approach, complements the previous three studies, achieving a similar amount of proteins and thus similar comprehensive coverage in the context of plant immunity. In immunity, very little is known about *A. thaliana* nuclear proteome including proteome rearrangement in the pattern triggered immunity (PTI). Only one previous short communication paper investigated the *A. thaliana* nuclear proteome in immunity after chitosan treatment (Fakih et al., 2016). This project investigated for the first time the *A. thaliana* nuclear proteome in PTI after flg22 and nlp20 treatments. Moreover, first steps were taken to investigate protein import into the nucleus and dual targeting of candidate proteins trafficked from other organelles under PTI. This discovery proteomics project led to the identification / quantification of hundreds of proteins in the *A. thaliana* nucleus in PTI and will provide a beneficial in-depth insight in the field of plant nuclear research, plant immunity and protein trafficking including nucleus-organelles communication. This project was LC-MS-based and therefore provides potential candidates for further verification with orthogonal methods, which could be done through one or several future sub-projects.

In this work the nuclear proteome was investigated under PTI using flg22 and nlp20, both PAMPS that trigger the immune response in a distinct ways (Wan et al., 2019). Flg22 is one of the best studied PAMPs, a 22 amino acid N-terminal epitope of bacterial flagellin (Meindl et al., 2000). It is recognized by LRR-RLK flagellin sensing 2 (FLS2) (Gómez-Gómez and Boller, 2000). More recently, the 20 amino acid peptide (nlp20), which is a distinctive part of the necrosis and ethylene-inducing peptide 1 (Nep1) like protein (NLPs), was shown to trigger the PTI after its recognition with the LRR-RP RLP23 receptor (Albert et al., 2015, Böhm et al., 2014). In contrast to flg22, NLPs are found in bacteria, fungi and oomycete (Dong et al., 2012, Qutob et al., 2006, Oome and Van den Ackerveken, 2014). Plant immunity is a complex process that involves an array of proteins and small molecules, particularly phytohormone, signaling, large scale reprograming of transcription and ultimately proteome remodeling. Our previous study (Bassal et al., 2020) has shown that 16 h after flg22 treatment, PTI is fully induced and proteome remodeling is complete as opposed to shorter time points of 1 and 3 h when transcriptional processes are more predominant. For this reason, in the nuclei isolation approach a long exposure time of 16 h was chosen to ensure full penetrance of PTI to the proteome. This time point will capture both early and late processes including import of differentially expressed proteins into the nucleus because of continuous exposure to the PAMP and full elicitation of immunity over a prolonged period. Conversely, in the proximity labeling experiment the continuous exposure to the flg22 was not possible due to the use of plant leaves, experimental set up and interference with biotin treatment. Moreover, the use of cycloheximide for a very long period could induce a high stress response from the plants. Therefore, a shorter time period of 5 h of flg22 treatment was used. This exposure time was long enough to investigate the import and trafficking of proteins at an early time point in PTI and still long enough to capture proteome rearrangement in PTI.

4.1 Defining the *Arabidopsis thaliana* nuclear proteome

Organelle proteomes, including the proteome of the nucleus, are by definition smaller than the total cellular proteome which comprises around 10,000 proteins at any given time (Nagaraj et al., 2011). Their qualitative composition, however, may be substantially more dynamic because of biological context-dependent protein import and export. In the organelle isolation approach the use of protoplasts to isolate the nuclei was successful and produced high quality nuclei in a round and intact form (Figure 9). With this approach, we sought to gain a first impression of subcellular protein distribution on a large scale by use of the LC-MS technology. Retrograde signaling from the chloroplast to the nucleus in plant immunity is well known. In the nuclei

isolation experiment cells cultured in the dark were used without chloroplasts to avoid the inevitable co-purification of these organelles with nuclei. Thus, the first approach does not provide any insights into protein import from the chloroplast. The isolated nuclear fraction was of high purity and was enriched in histones and not in cytoplasmic or mitochondrial markers (Figure 10 C), so our results disclose a number of candidate proteins that may be imported into the nucleus from these locations upon elicitation of PTI by both flg22 and nlp20.

An approach based on implementing a cut off limit was used to differentiate between nuclear and non-nuclear proteins. Generally cytoplasmic markers such as actin and phosphoenolpyruvate carboxylase are abundant in the cytoplasm and not expected to be found in the nucleus (Garcia et al., 2010, Genenncher et al., 2016, Dalmadi et al., 2019). Therefore, the cytoplasmic markers can act as representatives for the non-nuclear protein contaminants when found in the nucleus and their nuclear protein fraction (Npf_n) can be used as cut off limit to curate the nuclear proteins identified in the nuclear fraction. Arguably the median Npf_n values of the cytoplasmic markers (used as a cut-off limit) and the mitochondrial markers were around 0.25, indicating a large portion of the proteins we considered as genuinely localized in the nucleus were more abundant in the cellular fraction. Nonetheless, we think this approach is valid, because the selected markers are explicitly not nuclear and a substantial portion of the nuclear proteome is not facultative, i.e. found in other compartments and potentially underlying some trafficking. Indeed, the purity of the curated nuclear proteins was verified and 89 % were either experimentally verified nuclear proteins or proteins annotated / predicted to be localized in the nucleus. This reveals the success of the procedure and experimental approach used in characterizing the nuclear proteome.

In the second approach used in this project, TurboID was employed to biotinylate nuclear proteins that can be purified and enriched on streptavidin beads without nuclei isolation. This circumvents the contamination associated with organelle isolation techniques. Additionally, this approach allowed an easy identification of nuclear proteins directly in adult *A. thaliana* rosette leaves. Isolation of nuclei and investigation of nuclear proteins directly from adult *A. thaliana* plants is very difficult and was only successful two times before (Bae et al., 2003, Bigeard et al., 2014) and the rest of the studies used either cell culture, protoplasts or seedlings. Proximity dependent biotinylation and successful enrichment of biotinylated proteins are based on strong affinity of biotin to streptavidin (Chen and Perrimon, 2017, Gingras et al., 2019, Roesli et al., 2006); as a result it is very difficult to elute the biotinylated proteins from the streptavidin beads. Therefore, on bead digestion is always the method of choice that adds a very crucial step to

differentiate between the true positive biotinylated proteins and the proteins that bound non-specifically to the beads. This is usually achieved by the use of controls (Col-0, wild type) and stringent washing steps. The method used in this approach was developed and optimized based on previous protocols (Branon et al., 2018, Mair et al., 2019) and allowed a large reduction in the non-specific binding when compared to (Mair et al., 2019) who firstly validated the TurboID and investigated the nuclear proteins in *A. thaliana* seedlings. Only 4 % of proteins identified in transgenic plant expressing TurboID construct were also identified in WT indicating that they were bound non-specifically. Whereas in (Mair et al., 2019) about 75 % of proteins identified in transgenic plant expressing TurboID construct were also identified in WT. The purity of the nuclear proteins was verified and 93 % of the nuclear proteins consisted of either experimentally verified nuclear proteins or proteins annotated / predicted to be localized in the nucleus. This underscores the high quality of the nuclear proteome.

As discussed above and as mentioned in the results sections, both approaches produced a high quality nuclear proteome with high purity. In the organelle isolation approach, an arbitrary cut off limit was used which was shown to be efficient in curating the nuclear proteome, but still I can not claim the complete absence of inevitable contaminants associated with the nuclear isolation method. Proximity-dependent labeling approach successfully circumvented the contamination associated with organelle isolation techniques. Nevertheless, there might be a very small possibility of false nuclear identification with this approach. This could happen by labeling of proteins in cytosol by TurboID directly after translation before entering the nucleus. I believe that this will have a negligible negative effect on the quality of nuclear proteome for the following reasons. Firstly, TurboID-YFP-NLS is localized to the nucleus after translation. Inside the nucleus, it has enough time required for labeling the nuclear proteins with biotin. As a result, proteins inside the nucleus will be heavily biotinylated and more abundant than the proteins rapidly biotinylated outside the nucleus (on the way to the nucleus from the cytosol). Secondly, the biotinylated proteins will be enriched with streptavidin beads followed by a DDA scan strategy (TOP 10 abundancy). As a result, it is highly unlikely that the less abundant less biotinylated contaminants will be present in the final nuclear proteome.

Ten to twenty % (about 2000 proteins) of total cellular proteome are predicted to be nuclear at any given time (Narula et al., 2013). Both approaches used in this project identified a comparable number of proteins, which were also similar to amounts identified in the last three studies investigating the *A. thaliana* nuclear proteome comprehensively. This indicates the

comprehensiveness of both methods. Proteins identified and annotated to the nucleus in this project were involved in diverse nuclear functions. Transcription factors and transcriptional regulators control gene activity. Ribosomal proteins are part of the ribosomal biogenesis process (Watkins and Bohnsack, 2012, Turowski and Tollervey, 2015, Woolford and Baserga, 2013). Nucleotide binding proteins (DNA or RNA binding), modulate gene expression and kinases are components of signal transduction cascades that regulate gene expression (Hucho and Buchner, 1997). Translation initiation factors act as regulatory players in the translation process. WD40 proteins participate in various biological regulatory processes such as histone modifications, histone recognition and transcriptional regulation (Suganuma et al., 2008, Znaidi et al., 2004). Also, proteins involved in DNA repair, cell division and transport were identified.

Both methods used in this project were able to identify 231 proteins (nuclei isolation method) and 322 proteins (proximity labeling method) annotated to the transcription process in control conditions. The number of proteins, annotated to transcription identified in a recent paper addressing the *A. thaliana* nuclear proteome (again using DAVID) that employed the TurboID strategy for specific identification of nuclear proteins, were checked (Mair et al., 2019). It was 266, similar to the results in this project. Furthermore, I checked the number of TFs in the data of this project and in concatenated data from a number of previously published nuclear proteomics studies (Pendle et al., 2005, Calikowski et al., 2003, Bigeard et al., 2014, Sakamoto and Takagi, 2013, Chaki et al., 2015, Goto et al., 2019, Palm et al., 2016, Tang et al., 2020, Mair et al., 2019, Bae et al., 2003) that are found in PlantTFDB (Tian et al., 2019, Jin et al., 2016), a database of plant transcription factors. 138 TFs were identified in the data set of nuclei isolation method and 177 TFs were identified in the data set of proximity labeling method, whereas 231 were identified in the data from the other studies. Thus, this meta-analysis of TFs identified in this project and other *A. thaliana* nuclear proteomic studies shows that LC-MS measurements of the nuclear proteins, in both methods in this project, allowed deep insight into the transcription factor landscape in the nucleus which attests to the sequencing depth of both methods. However, this comparative analysis suggests a slight increase in the sensitivity of transcription factor detection in the second approach that employed the use of adult plants, proximity dependent labeling and LFQ as protein quantitative index.

Nuclear envelope proteins were also identified in this study. In *A. thaliana* 44 proteins were reported to be localized in the nuclear envelope, inner and outer membrane based on FP experimental evidence in SUBA4. 16 of these (36%) were identified in the nucleus at control conditions of the nuclei isolation experiment and 10 of these (23%) were identified in the nucleus at control conditions of the proximity labeling experiment. 692 proteins were found in

SUBA4 with the same localization based on FP and MS experimental evidence, this study identified 350 of these (51%) in control conditions of the nuclei isolation experiment and 130 (19%) in control conditions of the proximity labeling experiment. The lower numbers of NE proteins identified in the second approach could be explained by the fact that in the proximity labeling method, the TurboID fusion protein enters the nucleus and stays within to label the nuclear proteins. It will only have a chance to label NE proteins during its´s actual import and as a result, many NE proteins could be missed.

4.2 LC-MS-based analysis of dual targeted proteins

Proteins exert their functions in one or more organelles and it has been recognized that dual targeting of proteins in various developmental or stress response scenarios is an important but still understudied phenomenon especially in plants. Dual targeting was first identified 1995 (Creissen et al., 1995) but research in the last twenty years suggests that it may be a widespread event leading to the diversification of protein function (Krause and Krupinska, 2009, Sharma et al., 2018, Krupinska et al., 2020). Generally, dual localization can be classified into two main categories: dual targeting of newly synthesized proteins and relocalization/trafficking of mature proteins from one organelle to another (Krause and Krupinska, 2009, Krupinska et al., 2020) Query of the SUBA4 database with our list of nuclear proteins in control conditions (Table 14) identified several hundred proteins with known alternative sub-cellular localization possibly implicating them as dual targeted. Protein dual targeting has been studied mostly using FP based approaches and our list presents candidates that could be investigated further for confirmation.

In this study, the proteins import to the nucleus from cytosol or other organelles (dual targeted/trafficked) under PTI were also investigated. Several potential candidate proteins were identified and were annotated or known to be in the plasma membrane, cytosol, mitochondria, chloroplast and other organelles. Retrograde signaling from chloroplast and mitochondria is known as the back flow of instructions from these organelles to the nucleus reporting their functional state including the transfer of proteins (Bräutigam et al., 2007, Bobik and Burch-Smith, 2015). This process is known also to be engaged in the cellular response to biotic and abiotic stresses (Bobik and Burch-Smith, 2015, Crawford et al., 2018, Kmiecik et al., 2016). Furthermore, nucleocytoplasmic shuttling is a known biological process (Fu et al., 2018, Liu and Coaker, 2008). Recently, trafficking of proteins between the plasma membrane and the nucleus was shown to be frequent (Zheng and Jiang, 2022). Generally, protein transport into the nucleus

is controlled by different mechanisms. Proteins smaller than 40-60 kDa diffuse in a passive manner but larger proteins need to be recognized by the nuclear transport receptors which bind the nuclear localization signals (NLS) on those proteins and facilitate import (Mohr et al., 2009, Weis, 2003, Timney et al., 2016, Wang and Brattain, 2007, Chook and Suel, 2011, Christie et al., 2016, Grossman et al., 2012, Tamura and Hara-Nishimura, 2014). In addition, alternative mechanisms were also investigated (Imamoto and Kose, 2012, Guinez et al., 2005). In the nuclei isolation experiment two lists of proteins were identified as potential candidates for import under both treatment conditions (157 proteins for flg22 condition and 73 proteins for nlp20 condition) (Figure 12). In the list of flg22 challenge (Supplementary table 2), approx.33% of the proteins were predicted to have an NLS. Moreover, the molecular weights of the 71% of the remaining proteins in this list were less than 40-60 kDa, and thus below the size-exclusion limit of nuclear pore complexes. These latter proteins might enter the nucleus by diffusion and even if not possieeing an NLS. Similarly, in the corresponding list for nlp20-challenge (Supplementary table 2), approx. 33% of the proteins were predicted to have an NLS and 88% of the rest of the proteins in this list were less than 40-60 kDa. In the proximity labeling experiment a list of 21 proteins were identified as potential candidates for import from cytosol or other organelles under flg22 treatment (Figure 22), (Supplementary table 7). Approx. half of these proteins were predicted to have NLS. Moreover, the molecular weights of the 70% of the remaining proteins in this list were less than 40-60 kDa. Therefore, the majority of proteins we determined as nuclear meet the requirements for transport into the nucleus.

Proteins function in a specific manner, depending on time and location. Therefore, the designated proteins needed to be activated and recruited to certain subcellular locations only when required (Michaelson et al., 2008, Wiatrowski and Carlson, 2003). The activation could be through a signal transduction cascade that activates the protein by post translational modifications (Di Ventura and Kuhlman, 2016). In addition, nucleoporins regulate selectively the passage of certain stress-sensible proteins (Yang et al., 2017) by undergoing conformational changes upon receptor activation and allowing transport of specific macromolecules (Gu et al., 2016). In this work the nuclear import was investigated under flg22 and nlp20 stimulus, implying that the flg22 and nlp20 induced responses could directly or indirectly regulate nuclear import of the selected sets of proteins by one of the above-mentioned mechanisms.

In the second approach, cycloheximide was used to differentiate between the *de novo* (newly) synthesized proteins and subsequently imported into the nucleus and proteins that were already present and were more likely trafficked to the nucleus. 21 proteins were absent in nucleus under water treated control condition (W), were present in flg22 treated condition (F) and were present

in flg22 + cycloheximide treated condition (FC) (Figure 22) indicating that these sets of candidate proteins could be trafficked to the nucleus from the cytosol or some other organelle upon elicitation of immunity. Moreover, the results also showed that 90 proteins were absent in the nucleus under the water treated control condition (W) and were present in the nucleus under the cycloheximide treated condition (C). This indicates that the 90 proteins were possibly imported into the nucleus and could be trafficked from the cytosol or other organelles after the cycloheximide exerted stress and suggests a strong effect of cycloheximide on subcellular localization of proteins. The change in subcellular localization of proteins after cycloheximide treatment was also reported before for example, cycloheximide induced the internalization of EGF receptor and StSUT transporter (Oksvold et al., 2012, Garg et al., 2020). Interestingly, most of the 21 proteins absent in the nucleus in (W) and identified in (F) and (FC), were also common to the set of 90 proteins absent in (W) and present in (C). These proteins were present in F, FC and C. Therefore, I speculate that these proteins could be trafficking to the nucleus under biotic stress (PTI), but could also be trafficking to the nucleus generally under stress.

In this project, nuclear protein import was studied in PTI and several putative candidate proteins were identified for example: 51 kDa subunit of complex I and NAD-dependent malic enzyme 2 (2 mitochondrial proteins), calcium sensing receptor (chloroplast/mitochondrial protein) and Clathrin heavy chain 2 (Golgi/plasma membrane protein).

The 51 kDa subunit of complex I and NAD-dependent malic enzyme 2

In the nuclei isolation experiment the possible re-localization of these two proteins was investigated (see results section under 3.1.2). Most mitochondrial proteins contain transit peptides in their primary structure. Transit peptides are usually removed by mitochondrial processing peptidases (MPP) following import into mitochondria (Hawlitschek et al., 1988). 51 kDa subunit of complex I and NAD-dependent malic enzyme 2 are mitochondrial proteins that were previously reported also in the nucleus (Palm et al., 2016, Iglesias et al., 2013). 51 kDa subunit of complex I was identified as a substrate for the mitochondrial localized peptidase ICP55, which is a secondary processing peptidase that removes phenylalanine (F) from its MPP-processed form (Carrie et al., 2015). Phenylalanine was the first amino acid preceding the non-tryptic N-terminal peptide we identified in our MS results by way of no enzyme specificity search (Table 9). This may explain the non-tryptic cleavage site, giving a hint of possible primary processing of the protein's transit peptide by MPP followed by secondary removal of phenylalanine in the mitochondria before trafficking to the nucleus. Interestingly, phenylalanine

was again the first amino acid before the identified non-tryptic peptide for the NAD-dependent malic enzyme 2. In addition, ICP55 has a general consensus motif of RX (<u>F/Y/I/L</u>) (S/A) (S/T) where it removes the amino acids F, Y, I or L (Carrie et al., 2015). As shown in Table 9, the NAD-dependent malic enzyme 2 contains the ICP55 processing motif RC (<u>F</u>) (S) (T). This suggest that NAD-dependent malic enzyme 2 could also be processed by ICP55 after removal of the transit peptide by MPP. These MS findings require further verification by orthogonal methods.

<u>Calcium sensing receptor (CaS)</u>

Calcium sensing receptor is a non-EF hand calcium binding protein and is localized to the plasma membrane, chloroplast and mitochondria (Carrie et al., 2009b, Weinl et al., 2008, Han et al., 2003). CaS was reported in the thylakoid membrane to participate in cytoplasmic calcium regulation and is necessary in extracellular calcium induced stomatal closure (Nomura et al., 2008, Weinl et al., 2008, Vainonen et al., 2008). Previously a study reported that CaS, after PAMP stimulation, could regulate the transcription of defense genes through O_2 mediated (ROS) retrograde signaling (Nomura et al., 2012). In this thesis, the CaS was localized in the nucleus after flg22 treatment and could be shuttling to the nucleus from other organelles for example the chloroplast (Table 12). Several process have been reported to be dependent on nuclear calcium and nuclear localized calcium binding proteins (Charpentier, 2018). Additionally, nuclear and cytosolic calcium activate transcription in a distinct way (Hardingham et al., 1997) and have independent dynamics (Huang et al., 2017). Therefore, I speculate that CaS could be trafficked to the nucleus in PTI and involved in transcriptional programming by regulating nuclear calcium level.

<u>Clathrin heavy chain 2 (CHC2)</u>

Clathrin heavy chain 2 is a part of the clathrin coat protein complex localizing to both plasma membrane and trans Golgi network. It is engaged in clathrin-mediated endocytosis (CME) and post Golgi trafficking (Konopka et al., 2008, Ito et al., 2012, Dhonukshe et al., 2007, Castillon et al., 2018, Martzoukou et al., 2018). In the context of plant immunity, CHC2 plays a vital role in the CME of activated FLS2 receptors after flg22 perception (Mbengue et al., 2016, Ortiz-Morea et al., 2016). Impairment of internalization of the FLS2 receptor changes the defense response against infection (Robatzek et al., 2006, Spallek et al., 2013). Moreover, CHC2 variants were reported to interfere with auxin transporter internalization and CME dependent endocytotic

trafficking, also interfered with response to salicylic acid (Kitakura et al., 2011, Du et al., 2013). Recently, a study in *A. thaliana* reported possible communication between the CME (plasma membrane) and post Golgi trafficking (TGN) (Yan et al., 2021). Results from this thesis suggest that CHC2 may be a possible candidate for cross talk between the plasma membrane/Golgi and the nucleus after flg22 treatment by trafficking from these organelles to the nucleus (Table 12). In humans, monomeric CHC was reported to be localized in the nucleus and to activate the P53 transcription factor (Enari et al., 2006, Ohmori et al., 2008). P53 regulates many genes responsible for DNA repair, growth arrest and homologous recombination indicating distinct role of CHC in the nucleus (Seo et al., 2002, Bourdon et al., 2003, Linke et al., 2003). Therefore, I would surmise that CHC2 possible function in the nucleus under PTI could be regulating transcription factors involved directly or in directly in plant immune response.

4.3 Quantitative changes of proteome upon pattern triggered immunity

In this project, the nuclear proteomes were investigated for quantitative changes in protein abundance in between the control condition and the elicited PTI conditions. 115 proteins showed statistically significant changes in their abundance upon elicitation of one or both forms of PTI when compared to control. I will focus in the discussion on two proteins families: ribosomal proteins and prohibitin proteins and some proteins such as, phospholipase D alpha 1, WRKY transcription factors, Calcium binding EF-hand family protein (CML40) and Glutathione S-transferase (GSTL3).

Ribosomal proteins

The ribosomes are the cellular machinery required for the process of protein synthesis. The maturation of the ribosome is required for its function, the process of ribosome maturation is called ribosome biogenesis. Ribosome biogenesis involves association of ribosomal proteins with rRNA to constitute the ribosomal subunits (Thomson et al., 2013, Sáez-Vásquez and Delseny, 2019). Primary steps of ribosome biogenesis exist in the nucleus before exportation to the cytoplasm (Brown and Shaw, 1998, Stępiński, 2014, Henras et al., 2015, Woolford and Baserga, 2013). In addition to their function in ribosome biogenesis and protein synthesis ribosomal proteins have various extra-ribosomal functions (Zhou et al., 2015) for example, transcription regulation and histone binding in the nucleus (Denmat et al., 1994, Dieci et al., 2009, Tchórzewski et al., 1999, Tu et al., 2011, Ni et al., 2006). Therefore, ribosomal proteins have been identified in the nucleus in many studies for instance in *A. thaliana* (Pendle et al., 2005, Chaki et al., 2015, Palm et al., 2016). In this work ribosomal proteins were also identified

in the nuclear proteome (Table 10), 19 of them had a significant change in their abundance between control and elicited immunity. Interestingly, these ribosomal proteins showed different abundance in the nucleus when treated with the two elicitors flg22 and nlp20 as mentioned in the results. These elicitor specific changes in abundance suggest that ribosomal proteins do not act similarly and have different functions in the two types of PTI. Accordingly, we can speculate that the ribosomal proteins with increased abundance in the nucleus under flg22 and nlp20 compared to control (Ten proteins: five proteins specific to nlp20, two proteins specific to flg22 and three proteins for both flg22/nlp20) play an active role in the nucleus during the immune response under different stimuli in *A. thaliana*. On the other hand, the ribosomal proteins with decreased abundance in the nucleus under flg22 and nlp20 compared to control (9 proteins) have a repressed function in the nucleus during the *A. thaliana* immune response. In a previous study the ribosomal protein transcripts were investigated in *Vanilla planifolia* when infected with *Fusarium oxysporum* (Solano de la Cruz et al., 2019). Seven ribosomal proteins showed an increase in abundance in the nucleus after elicited immunity in our study and also showed an increase in their transcript expression patterns in *Vanilla* after two days of *Fusarium* infection. These seven protein families are: ribosomal protein L14p/L23e family protein, ribosomal L29 family protein, ribosomal protein L13 family protein, ribosomal protein L24e family protein, ribosomal protein L17 family protein, ribosomal protein S4 (RPS4A) family protein and ribosomal protein S5 domain 2-like superfamily protein. In addition, ribosomal protein L24e family protein was also detected exclusively in the nucleus of the *cerk1* background in *A. thaliana* after chitosan treatment (triggering a MAMP-like response). The authors also observed that the ribosomal proteins were overrepresented after chitosan treatment (Fakih et al., 2016). This suggest that these seven ribosomal proteins have distinct functions in plant immunity in different plants elicited by different pathogens and promiscuity of ribosomal proteins in ribosome assembly is known. This functional promiscuity is reflected by the different protein interactions undergone by the ribosomal proteins in the two PTI scenarios (Figure 13 C).

Prohibitins

Prohibitins are group of conserved proteins in eukaryotes including plants (Van Aken et al., 2010). They were reported to have several functions as scaffold proteins in mitochondrial biogenesis and immunity (Yu, 2019). Prohibitins participate in plant defense response and in protection against stress, for example: they are involved in the rice defense response against fungi (Takahashi et al., 1999, Takahashi et al., 2003). PHB1 and PHB2 are localized in the mitochondria and participate in its biogenesis and in the plant response to stress in *Nicotiana*

benthamiana (Ahn et al., 2006) and PHB3 is additionally localized to the chloroplast where it regulates the production of salicylic acid under UV and biotic stress in *A. thaliana* (Seguel et al., 2018). Besides their localization in the mitochondria and chloroplasts, prohibitins have also been reported in the nucleus and act as transcription regulators in eukaryotes (Huang et al., 2019, Mishra et al., 2006, Thuaud et al., 2013, Peng et al., 2015). In addition, PHB3 were localized by FP in the nucleus in *A. thaliana* (Pendle et al., 2005, Christians and Larsen, 2007, Huang et al., 2019*)* and possible shuttling between mitochondria and nucleus were also suggested (Yu, 2019). In *A thaliana* five prohibitins are expressed (PHB1, PHB2, PHB3, PHB4 and PHB6) (Van Aken et al., 2007) and all of them were identified in the nuclear proteome with increased abundance following treatment with both flg22 and nlp20 (Table 10). This indicates that the prohibitin family plays a role in the *A. thaliana* defense response in the nucleus. In previous studies, PHB2 was detected exclusively in the nucleus of *cerk1 A. thaliana* plant after chitosan treatment (triggering a MAMP-like response) (Fakih et al., 2016) and prohibitin protein was also identified in the nucleus of *Solanum lycopersicum* with increased abundance after 24 h infection with *Phytophthora capsici* compared to non-infected plants (Howden et al., 2017). The results of these two studies support our findings of a probable role of prohibitins in the nucleus during the plant immune response. In addition, all five prohibitins interacted with each other and Cytochrome C (Figure 13 B), another mitochondrial protein whose abundance also increased in the nucleus in PTI. Cytochrome C has been shown to have functions in the nucleus such as DNA damage repair and interaction with histone proteins (Gonzalez-Arzola et al., 2019) in addition to its well-known function in the mitochondrial respiratory chain. It is therefore tempting to speculate, that the prohibitins may act as a scaffold to traffic Cytochrome C from the mitochondrion to the nucleus in PTI.

<u>Phospholipase D alpha 1</u>

Phospholipase D alpha 1 (PLD) protein was identified in the nuclear proteome with increased abundance after treatment with both flg22 and nlp20 when compared to control indicating a possible role in the immune response within the nucleus (Table 10). Generally, PLD1 is involved in many plant cellular process including the plant response to wounding, salt stress and plant-microbial interactions (Wang, 2005, Bargmann and Munnik, 2006, Hong et al., 2010, Andersson et al., 2006, Zhao et al., 2013, Young et al., 1996). Additionally, an early study reported that PLD1 was localized to the nucleus in mammalian cell lines and had a possible function in mRNA processing (Jang and Min do, 2011). Similarly in *A. thaliana* PLD alpha and PLD gamma were also localized to the nucleus (Fan et al., 1999). Therefore, my findings are supported by

previous studies that link PLD to the plant defense response against microbes as well as its nuclear localization. Recently, a research group reported the interaction between PLD alpha 1 and MPK3 (mitogen-activated protein kinase 3) in regulating salt stress in *A. thaliana* (Vadovic et al., 2019). Mitogen-activated protein kinases play a significant role in plant defense response to biotic stress (Colcombet and Hirt, 2008, CristinaRodriguez et al., 2010, Šamajová et al., 2013, Latrasse et al., 2017) including MPK3 (Galletti et al., 2011) which was also localized to the nucleus (Vadovic et al., 2019). Hence, I can speculate on the possibility of a similar regulatory role of PLD alpha1 within the nucleus in the plant response to biotic stress upon interaction with MAPK or MPK3. Interestingly, Phospholipase D gamma 3 was identified in the nuclear proteome with decreased abundance after flg22/nlp20 treatment when compared to control, implying that different PLDs could have differing roles in nuclear associated processes in plant immunity.

WRKY transcription factors

WRKY transcription factors play a crucial role in regulating the transcription process in plant immunity. In the proximity labeling experiment WRKY 40 was identified in the nuclear proteome with increased abundance after treatment with flg22 when compared to control (Table 13 and Figure 23) indicating a possible active role within the nucleus in the *A. thaliana* immune response. This finding was supported by previous studies which reported that WRKY 40 plays an active role in the *A. thaliana* defense response against *Pseudomonas syringae* DC3000 either as a positive (Schön et al., 2013) or negative regulator (Lozano-Durán et al., 2013). Moreover, WRKY 40 play a role in coordinating the expression of nuclear encoded mitochondrial and chloroplast proteins in stress (Van Aken et al., 2013).

WRKY 72 is essential for basal defense in *A. thaliana* against root knot nematodes and the oomycete *Hyaloperonospora arabidopsidis*. Also, it is required in the basal defense against *Pseudomonas syringae* in tomato but not in *A. thaliana* (Bhattarai et al., 2010). Similarly in this project, WRKY 72 was identified with increased abundance in the nucleus after flg22 treatment (Table 13 and Figure 23) pointing out that, WRKY 72 possibly plays a role in *A. thaliana* defense against bacterial pathogens in PTI.

Calcium binding EF-hand family protein (CML40)

In 2017 a bioinformatics study showed that *A. thaliana* NPR1 and CML40 (calmodulin like protein 40) had high functional similarity (Yocgo et al., 2017). NPR1 is an essential regulator in plant immunity (Cao et al., 1997, Wu et al., 2012). More recently in 2019, an *in silico* study

reported a possible involvement of CML40 in the defense response against pathogens in plants (Tortosa et al., 2019). As shown in (Table 13 and Figure 23) CML40 was identified in the nuclear proteome with increased abundance after treatment with flg22 when compared to control. This result confirms the previous connections between the CML40 and immune response in plants and suggests a possible role of CML40 within the nucleus in PTI possibly by regulating nuclear calcium concentrations and directly or indirectly regulating transcription. In the same manner, two calcium binding proteins were identified in the nucleus of Tomato (*Solanum lycopersicum*) with increased abundance after 24 h infection with *Phytophthora capsici* compared to non-infected plants (Howden et al., 2017).

Glutathione S-transferase (GSTL3)

Plant GSTs are multifunctional proteins which are localized mostly in the cytosol and make up about 2% of soluble proteins (Pascal and Scalla, 1999, Gullner et al., 2018). They were reported to have multiple functions such as detoxification of toxins (Dixon et al., 1998, Schröder et al., 2007), auxin transport (Bilang and Sturm, 1995, Droog et al., 1995) and metabolism of anthocyanins (Marrs, 1996). In plant immunity several GSTs were involved in plant-pathogen interaction and antimicrobial resistance as reviewed by (Gullner et al., 2018). Besides their localization in the cytosol, some GSTs such as GSTT3L, GSTU12, GSTU19 and GSTF2 were also localized in the nucleus in plants and had a possible role in transcription regulation or DNA binding and nucleic acid repair (Dixon et al., 2009). In this project, GSTL3 was reported in the nucleus with increased abundance after flg22 treatment (Table 13 and Figure 23) indicating an active role in *A. thaliana* immune response within the nucleus. Recently a study identified a GST protein exclusively in the nucleus of PVA-infected potato leaves (Rajamaki et al., 2020) which support my finding of a probable role of GSTs including GSTL3 in the nucleus during the plant immune response.

Overall the experiments show a multitude of proteins associating with the nucleus upon PTI challenge, including proteins with diverse predicted physiological roles. This project will provide a beneficial in-depth insight in the field of plant nuclear research, plant immunity and protein trafficking including nucleus-organelles communication. The two approaches produced high quality catalogs of about 3000 nuclear proteins under control and PTI conditions including newly identified nuclear proteins. The measurement covered low abundant proteins including more than 200 transcription factors and transcriptional co-activators. First steps were taken to investigate protein import into the nucleus and dual targeted candidate proteins trafficked from

other organelles under PTI and found 178 and 73 proteins to possibly be imported into the nucleus upon stimulus with flg22 and nlp20, respectively. I also disclosed a list of several hundred of putatively dual targeted proteins under control (normal uninduced) conditions including proteins not yet found for further studies. In addition, the abundance of 115 proteins changed significantly in the nucleus following elicitation of immunity. The data will serve as a basis for further experiments aiming at characterizing the new candidates in more detail.

5. Outlook

This discovery proteomics project produced hundreds of protein candidates with potential functional roles in *Arabidopsis thaliana* nucleus under pattern triggered immunity (PTI) and will provide a beneficial in-depth insight in the field of plant nuclear research, plant immunity and protein trafficking including nucleus-organelles communication. However, this project is LC-MS-based and the results require further verification with orthogonal methods. An alternative confirmatory approach for subcellular localization of proteins is usually based on imaging the organelles using fluorescent proteins tagging (FP). Therefore, I would suggest one or several future sub-projects employing FP confirmatory methods. Nevertheless, this technique has also limitations: (1) the import of small sized proteins by passive diffusion to the nucleus could be impaired by the additional mass of FP tags. (2) The location of FP tags either C-terminal, N-terminal or in the middle could interfere with the import mechanism of the proteins. (3) The protein under investigation is usually overexpressed and is not reflecting the natural situation. These limitations could hinder the localization of the proteins with FP methods making it an exhausting and long process. As an example, using more than one FP-tagged construct per protein could be required to confirm the subcellular localization.

6. Summary

In spite of the vital role of the nucleus, the plant nuclear proteome remained understudied compared to the mammalian and human nuclear proteomes. Moreover, only few studies have investigated the core nuclear proteome of *Arabidopsis thaliana* comprehensively by LC-MS. In the context of plant immunity, very little is known about the *A. thaliana* nuclear proteome including proteome rearrangement in pattern triggered immunity (PTI). The communication and dynamics of plant cell organelles and how they coordinate their functions by trafficking of molecules and proteins, is a current focus of many research groups including my research training group.

This study aimed to use LC-MS-based discovery proteomic approaches to: (1) identify the *A. thaliana* nuclear proteome, (2) investigate the nuclear proteome under biotic stress stimuli, (3) investigate nuclear import including dual targeted proteins and trafficking from other compartments. In order to achieve the aim two main approaches were used. The first one was based on a nuclei isolation technique from *A. thaliana* cell culture, followed by devising an enrichment score to assess the purity of the preparation and curate the nuclear proteome. The second one utilized proximity dependent labeling to circumvent the inevitable contaminations. Both of them were used to investigate the nuclear proteome in the frame of PTI.

This discovery proteomics project led to the identification / quantification of hundreds of proteins in the *A. thaliana* nucleus under pattern triggered immunity (PTI) and will provide a beneficial in-depth insight in the field of plant nuclear research, plant immunity and protein trafficking including nucleus-organelles communication. The two approaches produced high quality catalogs of about 3000 nuclear proteins under control and PTI conditions including newly identified nuclear proteins. The measurement covered low abundant proteins including more than 200 transcription factors and transcriptional co-activators. First steps were taken to investigate protein import into the nucleus and dual targeted candidate proteins trafficked from other organelles under PTI and found 178 and 73 proteins to possibly be imported into the nucleus upon stimulus with flg22 and nlp20, respectively. I also disclosed a list of several hundred of putatively dual targeted proteins under control (normal uninduced) conditions including proteins not yet found for further studies. In addition, the abundance of 115 proteins changed significantly in the nucleus following elicitation of immunity. These results suggests a functional role of hundreds of proteins in the nucleus during PTI.

7. References

ABUKHALAF, M. 2018. Isolation of Nuclei for LC-MS Measurement of Nuclear Proteome. *Master thesis at Martin Luther University of Halle Wittenberg*.

ADE, J., DEYOUNG, B. J., GOLSTEIN, C. & INNES, R. W. 2007. Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proceedings of the National Academy of Sciences,* 104**,** 2531-2536.

AEBERSOLD, R. & CRAVATT, B. F. 2002. Proteomics – advances, applications and the challenges that remain. *Trends in Biotechnology,* 20**,** s1-s2.

AEBERSOLD, R. & MANN, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422**,** 198-207.

AGRAWAL, G. K., BOURGUIGNON, J., ROLLAND, N., EPHRITIKHINE, G., FERRO, M., JAQUINOD, M., ALEXIOU, K. G., CHARDOT, T., CHAKRABORTY, N. & JOLIVET, P. 2011. Plant organelle proteomics: collaborating for optimal cell function. *Mass Spectrometry Reviews,* 30**,** 772-853.

AHN, C. S., LEE, J. H., REUM HWANG, A., KIM, W. T. & PAI, H. S. 2006. Prohibitin is involved in mitochondrial biogenesis in plants. *Plant J,* 46**,** 658-67.

ALBERT, I., BÖHM, H., ALBERT, M., FEILER, C. E., IMKAMPE, J., WALLMEROTH, N., BRANCATO, C., RAAYMAKERS, T. M., OOME, S., ZHANG, H., KROL, E., GREFEN, C., GUST, A. A., CHAI, J., HEDRICH, R., VAN DEN ACKERVEKEN, G. & NÜRNBERGER, T. 2015. An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity. *Nature Plants,* 1**,** 15140.

ANDERSSON, M. X., KOURTCHENKO, O., DANGL, J. L., MACKEY, D. & ELLERSTRÖM, M. 2006. Phospholipase-dependent signalling during the AvrRpm1- and AvrRpt2-induced disease resistance responses in Arabidopsis thaliana. *The Plant journal : for cell and molecular biology,* 47**,** 947-959.

AYASH, M., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. 2021. LC-MS Based Draft Map of the Arabidopsis thaliana Nuclear Proteome and Protein Import in Pattern Triggered Immunity. *Frontiers in plant science,* 12**,** 744103-744103.

BAE, M. S., CHO, E. J., CHOI, E. Y. & PARK, O. K. 2003. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *The Plant Journal,* 36**,** 652-663.

BANTSCHEFF, M., LEMEER, S., SAVITSKI, M. M. & KUSTER, B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry,* 404**,** 939-965.

BARGMANN, B. O. R. & MUNNIK, T. 2006. The role of phospholipase D in plant stress responses. *Current Opinion in Plant Biology,* 9**,** 515-522.

BASSAL, M., ABUKHALAF, M., MAJOVSKY, P., THIEME, D., HERR, T., AYASH, M., TABASSUM, N., AL SHWEIKI, M. H. D. R., PROKSCH, C., HMEDAT, A., ZIEGLER, J., LEE, J., NEUMANN, S. & HOEHENWARTER, W. 2020. Reshaping of the Arabidopsis thaliana Proteome Landscape and Co-regulation of Proteins in Development and Immunity. *Molecular Plant,* 13**,** 1709-1732.

BATEMAN, N. W., GOULDING, S. P., SHULMAN, N. J., GADOK, A. K., SZUMLINSKI, K. K., MACCOSS, M. J. & WU, C. C. 2014. Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Molecular & Cellular Proteomics,* 13**,** 329-338.

BEDNAREK, P., PISLEWSKA-BEDNAREK, M., SVATOS, A., SCHNEIDER, B., DOUBSKY, J., MANSUROVA, M., HUMPHRY, M., CONSONNI, C., PANSTRUGA, R. & SANCHEZ-VALLET, A. 2009. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science,* 323**,** 101-106.

BEDNENKO, J., CINGOLANI, G. & GERACE, L. 2003. Nucleocytoplasmic transport: navigating the channel. *Traffic,* 4**,** 127-135.

BHATTACHARJEE, S., GARNER, C. M. & GASSMANN, W. 2013. New clues in the nucleus: transcriptional reprogramming in effector-triggered immunity. *Frontiers in plant science,* 4**,** 364.

BHATTARAI, K. K., ATAMIAN, H. S., KALOSHIAN, I. & EULGEM, T. 2010. WRKY72-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato R gene Mi-1. *The Plant Journal,* 63**,** 229-240.

BIGEARD, J., COLCOMBET, J. & HIRT, H. 2015. Signaling mechanisms in pattern-triggered immunity (PTI). *Molecular plant,* 8**,** 521-539.

BIGEARD, J., RAYAPURAM, N., BONHOMME, L., HIRT, H. & PFLIEGER, D. 2014. Proteomic and phosphoproteomic analyses of chromatin-associated proteins from Arabidopsis thaliana. *Proteomics,* 14**,** 2141-55.

BILANG, J. & STURM, A. 1995. Cloning and characterization of a glutathione S-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. *Plant Physiology,* 109**,** 253-260.

BILBAO, A., VARESIO, E., LUBAN, J., STRAMBIO-DE-CASTILLIA, C., HOPFGARTNER, G., MÜLLER, M. & LISACEK, F. 2015. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics,* 15**,** 964-980.

BOBIK, K. & BURCH-SMITH, T. M. 2015. Chloroplast signaling within, between and beyond cells. *Frontiers in plant science,* 6**,** 781.

BÖHM, H., ALBERT, I., OOME, S., RAAYMAKERS, T. M., VAN DEN ACKERVEKEN, G. & NÜRNBERGER, T. 2014. A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in Arabidopsis. *PLoS pathogens,* 10**,** e1004491.

BONDARENKO, P. V., CHELIUS, D. & SHALER, T. A. 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography– tandem mass spectrometry. *Analytical chemistry,* 74**,** 4741-4749.

BOSCH, J. A., CHEN, C. L. & PERRIMON, N. 2021. Proximity-dependent labeling methods for proteomic profiling in living cells: An update. *Wiley Interdisciplinary Reviews: Developmental Biology,* 10**,** e392.

BOUDSOCQ, M., WILLMANN, M. R., MCCORMACK, M., LEE, H., SHAN, L., HE, P., BUSH, J., CHENG, S.-H. & SHEEN, J. 2010. Differential innate immune signalling via Ca2+ sensor protein kinases. *Nature,* 464**,** 418-422.

BOURDON, J., DE LAURENZI, V., MELINO, G. & LANE, D. 2003. p53: 25 years of research and more questions to answer. *Cell Death & Differentiation,* 10**,** 397-399.

BOUTROT, F. & ZIPFEL, C. 2017. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annual review of phytopathology,* 55**,** 257-286.

BRANON, T. C., BOSCH, J. A., SANCHEZ, A. D., UDESHI, N. D., SVINKINA, T., CARR, S. A., FELDMAN, J. L., PERRIMON, N. & TING, A. Y. 2018. Efficient proximity labeling in living cells and organisms with TurboID. *Nature biotechnology,* 36**,** 880-887.

BRÄUTIGAM, K., DIETZEL, L. & PFANNSCHMIDT, T. 2007. Plastid-nucleus communication: anterograde and retrograde signalling in the development and function of plastids. *Cell and molecular biology of plastids.* Springer.

BROWN, J. W. S. & SHAW, P. J. 1998. Small Nucleolar RNAs and Pre-rRNA Processing in Plants. *The Plant Cell,* 10**,** 649.

CALIKOWSKI, T. T., MEULIA, T. & MEIER, I. 2003. A proteomic study of the arabidopsis nuclear matrix. *J Cell Biochem,* 90**,** 361-78.

CAO, H., GLAZEBROOK, J., CLARKE, J. D., VOLKO, S. & DONG, X. 1997. The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell,* 88**,** 57-63.

CARRIE, C., GIRAUD, E. & WHELAN, J. 2009a. Protein transport in organelles: dual targeting of proteins to mitochondria and chloroplasts. *The FEBS journal,* 276**,** 1187-1195.

CARRIE, C., KÜHN, K., MURCHA, M. W., DUNCAN, O., SMALL, I. D., O'TOOLE, N. & WHELAN, J. 2009b. Approaches to defining dual-targeted proteins in Arabidopsis. *The Plant Journal,* 57**,** 1128-1139.

CARRIE, C., VENNE, A. S., ZAHEDI, R. P. & SOLL, J. 2015. Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in Arabidopsis thaliana. *J Exp Bot,* 66**,** 2691-708.

CASTILLON, G. A., BURRIAT-COULERU, P., ABEGG, D., CRIADO SANTOS, N. & WATANABE, R. 2018. Clathrin and AP1 are required for apical sorting of glycosyl phosphatidyl inositol-anchored proteins in biosynthetic and recycling routes in Madin-Darby canine kidney cells. *Traffic,* 19**,** 215-228.

CATHERMAN, A. D., SKINNER, O. S. & KELLEHER, N. L. 2014. Top down proteomics: facts and perspectives. *Biochemical and biophysical research communications,* 445**,** 683-693.

CHAKI, M., SHEKARIESFAHLAN, A., AGEEVA, A., MENGEL, A., VON TOERNE, C., DURNER, J. & LINDERMAYR, C. 2015. Identification of nuclear target proteins for S-nitrosylation in pathogen-treated Arabidopsis thaliana cell cultures. *Plant Sci,* 238**,** 115-26.

CHARPENTIER, M. 2018. Calcium signals in the plant nucleus: origin and function. *Journal of Experimental Botany,* 69**,** 4165-4173.

CHELIUS, D. & BONDARENKO, P. V. 2002. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of proteome research,* 1**,** 317-323.

CHEN, C. L. & PERRIMON, N. 2017. Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdisciplinary Reviews: Developmental Biology,* 6**,** e272.

CHEN, Y. & LIU, L. 2019. Targeted proteomics. *Functional Proteomics.* Springer.

CHO, W. C. 2007. Proteomics technologies and challenges. *Genomics, proteomics & bioinformatics,* 5**,** 77-85.

CHOOK, Y. M. & SUEL, K. E. 2011. Nuclear import by karyopherin-betas: recognition and inhibition. *Biochim Biophys Acta,* 1813**,** 1593-606.

CHRISTIANS, M. J. & LARSEN, P. B. 2007. Mutational loss of the prohibitin AtPHB3 results in an extreme constitutive ethylene response phenotype coupled with partial loss of ethylene-inducible gene expression in Arabidopsis seedlings. *Journal of Experimental Botany,* 58**,** 2237-2248.

CHRISTIE, M., CHANG, C. W., RONA, G., SMITH, K. M., STEWART, A. G., TAKEDA, A. A., FONTES, M. R., STEWART, M., VERTESSY, B. G., FORWOOD, J. K. & KOBE, B. 2016. Structural Biology and Regulation of Protein Import into the Nucleus. *J Mol Biol,* 428**,** 2060-90.

CHRISTOPHER, J. A., STADLER, C., MARTIN, C. E., MORGENSTERN, M., PAN, Y., BETSINGER, C. N., RATTRAY, D. G., MAHDESSIAN, D., GINGRAS, A.-C. & WARSCHEID, B. 2021. Subcellular proteomics. *Nature Reviews Methods Primers,* 1**,** 1-24.

CISKA, M. & MORENO DÍAZ DE LA ESPINA, S. 2013. NMCP/LINC proteins: putative lamin analogs in plants? *Plant signaling & behavior,* 8**,** e26669.

COLCOMBET, J. & HIRT, H. 2008. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochemical Journal,* 413**,** 217-226.

CONLAN, B., STOLL, T., GORMAN, J. J., SAUR, I. & RATHJEN, J. P. 2018. Development of a rapid in planta BioID system as a probe for plasma membrane-associated immunity proteins. *Frontiers in plant science***,** 1882.

COOPER, B., CAMPBELL, K. B., FENG, J., GARRETT, W. M. & FREDERICK, R. 2011. Nuclear proteomic changes linked to soybean rust resistance. *Molecular BioSystems,* 7**,** 773-783.

COX, J., HEIN, M. Y., LUBER, C. A., PARON, I., NAGARAJ, N. & MANN, M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics,* 13**,** 2513-2526.

COX, J. & MANN, M. 2011. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry,* 80**,** 273-299.

CRAWFORD, T., LEHOTAI, N. & STRAND, Å. 2018. The role of retrograde signals during plant stress responses. *Journal of experimental botany,* 69**,** 2783-2795.

CREISSEN, G., REYNOLDS, H., XUE, Y. & MULLINEAUX, P. 1995. Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. *Plant J,* 8**,** 167-75.

CRISTINARODRIGUEZ, M., PETERSEN, M. & MUNDY, J. 2010. Mitogen-Activated Protein Kinase Signaling in Plants. *Annual Review of Plant Biology,* 61**,** 621-649.

DALMADI, A., GYULA, P., BALINT, J., SZITTYA, G. & HAVELDA, Z. 2019. AGO-unbound cytosolic pool of mature miRNAs in plant cells reveals a novel regulatory step at AGO1 loading. *Nucleic Acids Res,* 47**,** 9803-9817.

DANGL, J. L. & JONES, J. D. 2001. Plant pathogens and integrated defence responses to infection. *nature,* 411**,** 826-833.

DE BOER, E., RODRIGUEZ, P., BONTE, E., KRIJGSVELD, J., KATSANTONI, E., HECK, A., GROSVELD, F. & STROUBOULIS, J. 2003. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences,* 100**,** 7480-7485.

DENMAT, S. H.-L., SIPICZKI, M. & THURIAUX, P. 1994. Suppression of Yeast RNA Polymerase III Mutations by the URP2 Gene Encoding a Protein Homologous to the Mammalian Ribosomal Protein S20. *Journal of Molecular Biology,* 240**,** 1-7.

DESLANDES, L. & RIVAS, S. 2011. The plant cell nucleus: a true arena for the fight between plants and pathogens. *Plant Signaling & Behavior,* 6**,** 42-48.

DHONUKSHE, P., ANIENTO, F., HWANG, I., ROBINSON, D. G., MRAVEC, J., STIERHOF, Y.-D. & FRIML, J. 2007. Clathrin-mediated constitutive endocytosis of PIN auxin efflux carriers in Arabidopsis. *Current Biology,* 17**,** 520-527.

DI VENTURA, B. & KUHLMAN, B. 2016. Go in! Go out! Inducible control of nuclear localization. *Curr Opin Chem Biol,* 34**,** 62-71.

DIECI, G., RUOTOLO, R., BRAGLIA, P., CARLES, C., CARPENTIERI, A., AMORESANO, A. & OTTONELLO, S. 2009. Positive modulation of RNA polymerase III transcription by ribosomal proteins. *Biochemical and Biophysical Research Communications,* 379**,** 489-493.

DIXON, D. P., CUMMINS, I., COLE, D. J. & EDWARDS, R. 1998. Glutathione-mediated detoxification systems in plants. *Current opinion in plant biology,* 1**,** 258-266.

DIXON, D. P., HAWKINS, T., HUSSEY, P. J. & EDWARDS, R. 2009. Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. *Journal of experimental botany,* 60**,** 1207-1218.

DODDS, P. N., LAWRENCE, G. J., CATANZARITI, A.-M., TEH, T., WANG, C.-I., AYLIFFE, M. A., KOBE, B. & ELLIS, J. G. 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences,* 103**,** 8888-8893.

DONG, S., KONG, G., QUTOB, D., YU, X., TANG, J., KANG, J., DAI, T., WANG, H., GIJZEN, M. & WANG, Y. 2012. The NLP toxin family in Phytophthora sojae includes rapidly evolving groups that lack necrosis-inducing activity. *Molecular Plant-Microbe Interactions,* 25**,** 896-909.

DROOG, F. N., HOOYKAAS, P. J. & VAN DER ZAAL, B. J. 1995. 2, 4-Dichlorophenoxyacetic acid and related chlorinated compounds inhibit two auxin-regulated type-III tobacco glutathione S-transferases. *Plant physiology,* 107**,** 1139-1146.

DU, Y., TEJOS, R., BECK, M., HIMSCHOOT, E., LI, H., ROBATZEK, S., VANNESTE, S. & FRIML, J. 2013. Salicylic acid interferes with clathrin-mediated endocytic protein trafficking. *Proceedings of the National Academy of Sciences,* 110**,** 7946-7951.

DUNKLEY, T. P., WATSON, R., GRIFFIN, J. L., DUPREE, P. & LILLEY, K. S. 2004. Localization of organelle proteins by isotope tagging (LOPIT). *Molecular & Cellular Proteomics,* 3**,** 1128-1134.

EDMAN, P. & BEGG, G. 1967. A protein sequenator. *European Journal of Biochemistry.* Springer.

EDMAN, P., HÖGFELDT, E., SILLÉN, L. G. & KINELL, P.-O. 1950. Method for determination of the amino acid sequence in peptides. *Acta chem. scand,* 4**,** 283-293.

ENARI, M., OHMORI, K., KITABAYASHI, I. & TAYA, Y. 2006. Requirement of clathrin heavy chain for p53-mediated transcription. *Genes & development,* 20**,** 1087-1099.

ENG, J. K., MCCORMACK, A. L. & YATES, J. R. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the american society for mass spectrometry,* 5**,** 976-989.

ERHARDT, M., ADAMSKA, I. & FRANCO, O. L. 2010. Plant nuclear proteomics–inside the cell maestro. *The FEBS Journal,* 277**,** 3295-3307.

FAKIH, Z., AHMED, M. B., LETANNEUR, C. & GERMAIN, H. 2016. An unbiased nuclear proteomics approach reveals novel nuclear protein components that participates in MAMP-triggered immunity. *Plant Signal Behav,* 11**,** e1183087.

FAN, L., ZHENG, S., CUI, D. & WANG, X. 1999. Subcellular distribution and tissue expression of phospholipase Dalpha, Dbeta, and Dgamma in Arabidopsis. *Plant physiology,* 119**,** 1371-1378.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.

FERGUSON, P. L. & SMITH, R. D. 2003. Proteome Analysis by Mass Spectrometry. *Annual Review of Biophysics and Biomolecular Structure,* 32**,** 399-424.

FOLTA, K. M. & KAUFMAN, L. S. 2006. Isolation of Arabidopsis nuclei and measurement of gene transcription rates using nuclear run-on assays. *Nature protocols,* 1**,** 3094-3100.

FU, X., LIANG, C., LI, F., WANG, L., WU, X., LU, A., XIAO, G. & ZHANG, G. 2018. The rules and functions of nucleocytoplasmic shuttling proteins. *International journal of molecular sciences,* 19**,** 1445.

GALICHET, A., HOYEROVÁ, K., KAMÍNEK, M. & GRUISSEM, W. 2008. Farnesylation directs AtIPT3 subcellular localization and modulates cytokinin biosynthesis in Arabidopsis. *Plant Physiology,* 146**,** 1155-1164.

GALLETTI, R., FERRARI, S. & DE LORENZO, G. 2011. Arabidopsis MPK3 and MPK6 play different roles in basal and oligogalacturonide- or flagellin-induced resistance against Botrytis cinerea. *Plant Physiol,* 157**,** 804-14.

GALLIEN, S. & DOMON, B. 2015. Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods,* 81**,** 15-23.

GALLIEN, S., PETERMAN, S., KIYONAMI, R., SOUADY, J., DURIEZ, E., SCHOEN, A. & DOMON, B. 2012. Highly multiplexed targeted proteomics using precise control of peptide retention time. *Proteomics,* 12**,** 1122-1133.

GAOUAR, O. & GERMAIN, H. 2013. mRNA export: threading the needle. *Frontiers in plant science,* 4**,** 59.

GARCIA, A. V., BLANVILLAIN-BAUFUME, S., HUIBERS, R. P., WIERMER, M., LI, G., GOBBATO, E., RIETZ, S. & PARKER, J. E. 2010. Balanced nuclear and cytoplasmic activities of EDS1 are required for a complete plant innate immune response. *PLoS Pathog,* 6**,** e1000970.

GARG, V., HACKEL, A. & KÜHN, C. 2020. Subcellular targeting of plant sucrose transporters is affected by their oligomeric state. *Plants,* 9**,** 158.

GENENNCHER, B., WIRTHMUELLER, L., ROTH, C., KLENKE, M., MA, L., SHARON, A. & WIERMER, M. 2016. Nucleoporin-Regulated MAP Kinase Signaling in Immunity to a Necrotrophic Fungal Pathogen. *Plant Physiol,* 172**,** 1293-1305.

GIEGÉ, P., HEAZLEWOOD, J. L., ROESSNER-TUNALI, U., MILLAR, A. H., FERNIE, A. R., LEAVER, C. J. & SWEETLOVE, L. J. 2003. Enzymes of glycolysis are functionally associated with the mitochondrion in Arabidopsis cells. *The Plant Cell,* 15**,** 2140-2151.

GILHAM, P. T. 1970. RNA SEQUENCE ANALYSIS. *Annual Review of Biochemistry,* 39**,** 227-250.

GILLET, L. C., NAVARRO, P., TATE, S., RÖST, H., SELEVSEK, N., REITER, L., BONNER, R. & AEBERSOLD, R. 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics,* 11.

GINGRAS, A.-C., ABE, K. T. & RAUGHT, B. 2019. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Current opinion in chemical biology,* 48**,** 44-54.

GO, C. D., KNIGHT, J. D., RAJASEKHARAN, A., RATHOD, B., HESKETH, G. G., ABE, K. T., YOUN, J.-Y., SAMAVARCHI-TEHRANI, P., ZHANG, H. & ZHU, L. Y. 2021. A proximity-dependent biotinylation map of a human cell. *Nature,* 595**,** 120-124.

GOLDFARB, D. S., CORBETT, A. H., MASON, D. A., HARREMAN, M. T. & ADAM, S. A. 2004. Importin α: a multipurpose nuclear-transport receptor. *Trends in cell biology,* 14**,** 505-514.

GÓMEZ-GÓMEZ, L. & BOLLER, T. 2000. FLS2: An LRR Receptor–like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. *Molecular Cell,* 5**,** 1003-1011.

GONZALEZ-ARZOLA, K., VELAZQUEZ-CRUZ, A., GUERRA-CASTELLANO, A., CASADO-COMBRERAS, M. A., PEREZ-MEJIAS, G., DIAZ-QUINTANA, A., DIAZ-MORENO, I. & DE LA ROSA, M. A. 2019. New moonlighting functions of mitochondrial cytochrome c in the cytoplasm and nucleus. *FEBS Lett,* 593**,** 3101-3119.

GÖRLICH, D. 1998. Transport into and out of the cell nucleus. *The EMBO journal,* 17**,** 2721-2727.

GOTO, C., HASHIZUME, S., FUKAO, Y., HARA-NISHIMURA, I. & TAMURA, K. 2019. Comprehensive nuclear proteome of Arabidopsis obtained by sequential extraction. *Nucleus,* 10**,** 81-92.

GROSSMAN, E., MEDALIA, O. & ZWERGER, M. 2012. Functional architecture of the nuclear pore complex. *Annu Rev Biophys,* 41**,** 557-84.

GU, Y., ZEBELL, S. G., LIANG, Z., WANG, S., KANG, B. H. & DONG, X. 2016. Nuclear Pore Permeabilization Is a Convergent Signaling Event in Effector-Triggered Immunity. *Cell,* 166**,** 1526-1538 e11.

GUINEZ, C., MORELLE, W., MICHALSKI, J. C. & LEFEBVRE, T. 2005. O-GlcNAc glycosylation: a signal for the nuclear transport of cytosolic proteins? *Int J Biochem Cell Biol,* 37**,** 765-74.

GULLNER, G., KOMIVES, T., KIRÁLY, L. & SCHRÖDER, P. 2018. Glutathione S-transferase enzymes in plant-pathogen interactions. *Frontiers in plant science,* 9**,** 1836.

GUST, A. A., WILLMANN, R., DESAKI, Y., GRABHERR, H. M. & NÜRNBERGER, T. 2012. Plant LysM proteins: modules mediating symbiosis and immunity. *Trends in plant science,* 17**,** 495-502.

GYGI, S. P., RIST, B., GERBER, S. A., TURECEK, F., GELB, M. H. & AEBERSOLD, R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology,* 17**,** 994-999.

HAN, S., TANG, R., ANDERSON, L. K., WOERNER, T. E. & PEI, Z.-M. 2003. A cell surface receptor mediates extracellular Ca2+ sensing in guard cells. *Nature,* 425**,** 196-200.

HAN, X., ASLANIAN, A. & YATES III, J. R. 2008. Mass spectrometry for proteomics. *Current opinion in chemical biology,* 12**,** 483-490.

HARDINGHAM, G. E., CHAWLA, S., JOHNSON, C. M. & BADING, H. 1997. Distinct functions of nuclear and cytoplasmic calcium in the control of gene expression. *Nature,* 385**,** 260-265.

HAWLITSCHEK, G., SCHNEIDER, H., SCHMIDT, B., TROPSCHUG, M., HARTL, F.-U. & NEUPERT, W. 1988. Mitochondrial protein import: Identification of processing peptidase and of PEP, a processing enhancing protein. *Cell,* 53**,** 795-806.

HENRAS, A. K., PLISSON-CHASTANG, C., O'DONOHUE, M.-F., CHAKRABORTY, A. & GLEIZES, P.-E. 2015. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley interdisciplinary reviews. RNA,* 6**,** 225-242.

HENZEL, W. J., BILLECI, T. M., STULTS, J. T., WONG, S. C., GRIMLEY, C. & WATANABE, C. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences,* 90**,** 5011-5015.

HOEHENWARTER, W., THOMAS, M., NUKARINEN, E., EGELHOFER, V., RÖHRIG, H., WECKWERTH, W., CONRATH, U. & BECKERS, G. J. 2013. Identification of novel in vivo MAP kinase substrates in Arabidopsis thaliana through use of tandem metal oxide affinity chromatography. *Molecular & Cellular Proteomics,* 12**,** 369-380.

HONG, Y., ZHANG, W. & WANG, X. 2010. Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity. *Plant, cell & environment,* 33 4**,** 627-35.

HOOPER, C. M., CASTLEDEN, I. R., TANZ, S. K., ARYAMANESH, N. & MILLAR, A. H. 2017. SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic acids research,* 45**,** D1064-D1074.

HOWARD, B. E., HU, Q., BABAOGLU, A. C., CHANDRA, M., BORGHI, M., TAN, X., HE, L., WINTER-SEDEROFF, H., GASSMANN, W. & VERONESE, P. 2013. High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. *PLoS One,* 8**,** e74183.

HOWDEN, A. J. M., STAM, R., MARTINEZ HEREDIA, V., MOTION, G. B., TEN HAVE, S., HODGE, K., MARQUES MONTEIRO AMARO, T. M. & HUITEMA, E. 2017. Quantitative analysis of the tomato nuclear proteome during Phytophthora capsici infection unveils regulators of immunity. *New Phytol,* 215**,** 309-322.

HU, A., NOBLE, W. S. & WOLF-YADLIN, A. 2016. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research,* 5. [version 1; peer review: 3 approved]

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research,* 37**,** 1-13.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4**,** 44-57.

HUANG, F., LUO, J., NING, T., CAO, W., JIN, X., ZHAO, H., WANG, Y. & HAN, S. 2017. Cytosolic and nucleosolic calcium signaling in response to osmotic and salt stresses are independent of each other in roots of Arabidopsis seedlings. *Frontiers in plant science,* 8**,** 1648.

HUANG, R., SHU, S., LIU, M., WANG, C., JIANG, B., JIANG, J., YANG, C. & ZHANG, S. 2019. Nuclear Prohibitin3 Maintains Genome Integrity and Cell Proliferation in the Root Meristem through Minichromosome Maintenance 2. *Plant physiology,* 179**,** 1669-1691.

HUCHO, F. & BUCHNER, K. 1997. Signal Transduction and Protein Kinases: The Long Way from the Plasma Membrane into the Nucleus. *Naturwissenschaften,* 84**,** 281-290.

IGLESIAS, J., TRIGUEROS, M., ROJAS-TRIANA, M., FERNÁNDEZ, M., ALBAR, J. P., BUSTOS, R., PAZ-ARES, J. & RUBIO, V. 2013. Proteomics identifies ubiquitin–proteasome targets and new roles for chromatin-remodeling in the Arabidopsis response to phosphate starvation. *Journal of Proteomics,* 94**,** 1-22.

IMAMOTO, N. & KOSE, S. 2012. Heat-shock stress activates a novel nuclear import pathway mediated by Hikeshi. *Nucleus,* 3**,** 422-8.

ITO, E., FUJIMOTO, M., EBINE, K., UEMURA, T., UEDA, T. & NAKANO, A. 2012. Dynamic behavior of clathrin in Arabidopsis thaliana unveiled by live imaging. *The Plant Journal,* 69**,** 204-216.

JANG, Y. H. & MIN DO, S. 2011. Nuclear localization of phospholipase D1 mediates the activation of nuclear protein kinase C(alpha) and extracellular signal-regulated kinase signaling pathways. *J Biol Chem,* 286**,** 4680-9.

JEZ, J. M., TOPP, C. N., SMYTHERS, A. L. & HICKS, L. M. 2021. Mapping the plant proteome: tools for surveying coordinating pathways. *Emerging Topics in Life Sciences,* 5**,** 203-220.

JIN, J., TIAN, F., YANG, D.-C., MENG, Y.-Q., KONG, L., LUO, J. & GAO, G. 2016. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research,* 45**,** D1040-D1045.

JONES, J. D. & DANGL, J. L. 2006. The plant immune system. *nature,* 444**,** 323-329.

KADOTA, Y., SKLENAR, J., DERBYSHIRE, P., STRANSFELD, L., ASAI, S., NTOUKAKIS, V., JONES, J. D., SHIRASU, K., MENKE, F. & JONES, A. 2014. Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. *Molecular cell,* 54**,** 43-55.

KALTSCHMIDT, E. T. & WITTMANN, H. 1970. Ribosomal proteins. VII: Two-dimensional polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. *Analytical biochemistry,* 36**,** 401-412.

KARAS, M. & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry,* 60**,** 2299-2301.

KELLIE, J. F., TRAN, J. C., LEE, J. E., AHLF, D. R., THOMAS, H. M., NTAI, I., CATHERMAN, A. D., DURBIN, K. R., ZAMDBORG, L. & VELLAICHAMY, A. 2010. The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Molecular BioSystems,* 6**,** 1532-1539.

KEMINER, O. & PETERS, R. 1999. Permeability of single nuclear pores. *Biophysical journal,* 77**,** 217-228.

KHAN, M., YOUN, J.-Y., GINGRAS, A.-C., SUBRAMANIAM, R. & DESVEAUX, D. 2018. In planta proximity dependent biotin identification (BioID). *Scientific reports,* 8**,** 1-8.

KIM, D. I., JENSEN, S. C., NOBLE, K. A., KC, B., ROUX, K. H., MOTAMEDCHABOKI, K. & ROUX, K. J. 2016. An improved smaller biotin ligase for BioID proximity labeling. *Molecular biology of the cell,* 27**,** 1188-1196.

KITAKURA, S., VANNESTE, S., ROBERT, S., LÖFKE, C., TEICHMANN, T., TANAKA, H. & FRIML, J. 2011. Clathrin mediates endocytosis and polar distribution of PIN auxin transporters in Arabidopsis. *The Plant Cell,* 23**,** 1920-1931.

KMIECIK, P., LEONARDELLI, M. & TEIGE, M. 2016. Novel connections in plant organellar signalling link different stress responses and signalling pathways. *Journal of Experimental Botany,* 67**,** 3793-3807.

KOBS, G. 1998. Isolation of RNA from plant, yeast, and bacteria. *Promega notes,* 68**,** 28.

KONOPKA, C. A., BACKUES, S. K. & BEDNAREK, S. Y. 2008. Dynamics of Arabidopsis dynamin-related protein 1C and a clathrin light chain at the plasma membrane. *The Plant Cell,* 20**,** 1363-1380.

KRASNY, L. & HUANG, P. H. 2021. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Molecular omics,* 17**,** 29-42.

KRAUSE, K. & KRUPINSKA, K. 2009. Nuclear regulators with a second home in organelles. *Trends Plant Sci,* 14**,** 194-9.

KRUPINSKA, K., BLANCO, N. E., OETKE, S. & ZOTTINI, M. 2020. Genome communication in plants mediated by organelle–nucleus-located proteins. *Philosophical Transactions of the Royal Society B,* 375**,** 20190397.

KUMAR, R., KUMAR, A., SUBBA, P., GAYALI, S., BARUA, P., CHAKRABORTY, S. & CHAKRABORTY, N. 2014. Nuclear phosphoproteome of developing chickpea seedlings (Cicer arietinum L.) and protein-kinase interaction network. *Journal of Proteomics,* 105**,** 58-73.

KUTAY, U. & HETZER, M. W. 2008. Reorganization of the nuclear envelope during open mitosis. *Current opinion in cell biology,* 20**,** 669-677.

LAM, Y. W., LAMOND, A. I., MANN, M. & ANDERSEN, J. S. 2007. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Current Biology,* 17**,** 749-760.

LANGE, V., PICOTTI, P., DOMON, B. & AEBERSOLD, R. 2008. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology,* 4**,** 222.

LATRASSE, D., JÉGU, T., LI, H., DE ZELICOURT, A., RAYNAUD, C., LEGRAS, S., GUST, A., SAMAJOVA, O., VELUCHAMY, A., RAYAPURAM, N., RAMIREZ-PRADO, J. S., KULIKOVA, O., COLCOMBET, J., BIGEARD, J., GENOT, B., BISSELING, T., BENHAMED, M. & HIRT, H. 2017. MAPK-triggered chromatin reprogramming by histone deacetylase in plant innate immunity. *Genome biology,* 18**,** 131-131.

LEE, H. Y., KIM, E. G., JUNG, H. R., JUNG, J. W., KIM, H. B., CHO, J. W., KIM, K. M. & YI, E. C. 2019a. Refinements of LC-MS/MS spectral counting statistics improve quantification of low abundance proteins. *Scientific reports,* 9**,** 1-10.

LEE, M. H., JEON, H. S., KIM, S. H., CHUNG, J. H., ROPPOLO, D., LEE, H. J., CHO, H. J., TOBIMATSU, Y., RALPH, J. & PARK, O. K. 2019b. Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. *The EMBO journal,* 38**,** e101948.

LI, P., LI, J., WANG, L. & DI, L. J. 2017. Proximity labeling of interacting proteins: application of BioID as a discovery tool. *Proteomics,* 17**,** 1700002.

LIN, Q., ZHOU, Z., LUO, W., FANG, M., LI, M. & LI, H. 2017. Screening of proximal and interacting proteins in rice protoplasts by proximity-dependent biotinylation. *Frontiers in plant science,* 8**,** 749.

LINKE, S. P., SENGUPTA, S., KHABIE, N., JEFFRIES, B. A., BUCHHOP, S., MISKA, S., HENNING, W., PEDEUX, R., WANG, X. W. & HOFSETH, L. J. 2003. p53 interacts with hRAD51 and hRAD54, and directly modulates homologous recombination. *Cancer research,* 63**,** 2596-2605.

LIU, H., SADYGOV, R. G. & YATES, J. R. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical chemistry,* 76**,** 4193-4201.

LIU, J. & COAKER, G. 2008. Nuclear trafficking during plant innate immunity. *Molecular plant,* 1**,** 411-422.

LOO, R. R. O., HAYES, R., YANG, Y., HUNG, F., RAMACHANDRAN, P., KIM, N., GUNSALUS, R. & LOO, J. A. 2005. Top-down, bottom-up, and side-to-side proteomics with virtual 2-D gels. *International Journal of Mass Spectrometry,* 240**,** 317-325.

LOTT, K. & CINGOLANI, G. 2011. The importin β binding domain as a master regulator of nucleocytoplasmic transport. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1813**,** 1578-1592.

LOUREIRO, J., RODRIGUEZ, E., DOLEŽEL, J. & SANTOS, C. 2006. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Annals of Botany,* 98**,** 679-689.

LOUREIRO, J., RODRIGUEZ, E., DOLEŽEL, J. & SANTOS, C. 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of botany,* 100**,** 875-888.

LOZANO-DURÁN, R., MACHO, A. P., BOUTROT, F., SEGONZAC, C., SOMSSICH, I. E. & ZIPFEL, C. 2013. The transcriptional regulator BZR1 mediates trade-off between plant innate immunity and growth. *elife,* 2**,** e00983.

LU, D., WU, S., GAO, X., ZHANG, Y., SHAN, L. & HE, P. 2010. A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proceedings of the National Academy of Sciences,* 107**,** 496-501.

MA, J., GORYAYNOV, A., SARMA, A. & YANG, W. 2012. Self-regulated viscous channel in the nuclear pore complex. *Proceedings of the National Academy of Sciences,* 109**,** 7326-7331.

MACHO, A. P. & ZIPFEL, C. 2014. Plant PRRs and the activation of innate immune signaling. *Molecular cell,* 54**,** 263-272.

MACKO, V. & STEGEMANN, H. 1969. Mapping of potato proteins by combined electrofocusing and electrophoresis identification of varieties. *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie,* 350**,** 917-919.

MAIR, A., XU, S.-L., BRANON, T. C., TING, A. Y. & BERGMANN, D. C. 2019. Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. *Elife,* 8**,** e47864.

MAJOVSKY, P., NAUMANN, C., LEE, C.-W., LASSOWSKAT, I., TRUJILLO, M., DISSMEYER, N. & HOEHENWARTER, W. 2014. Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. *Journal of proteome research,* 13**,** 4246-4258.

MAKAROV, A., DENISOV, E., KHOLOMEEV, A., BALSCHUN, W., LANGE, O., STRUPAT, K. & HORNING, S. 2006. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry,* 78**,** 2113-2120.

MALINOVSKY, F. G., FANGEL, J. U. & WILLATS, W. G. 2014. The role of the cell wall in plant immunity. *Frontiers in plant science,* 5**,** 178.

MANN, M., HENDRICKSON, R. C. & PANDEY, A. 2001. Analysis of proteins and proteomes by mass spectrometry. *Annual review of biochemistry,* 70**,** 437-473.

MARFORI, M., MYNOTT, A., ELLIS, J. J., MEHDI, A. M., SAUNDERS, N. F., CURMI, P. M., FORWOOD, J. K., BODÉN, M. & KOBE, B. 2011. Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1813**,** 1562-1577.

MARRS, K. A. 1996. The functions and regulation of glutathione S-transferases in plants. *Annual review of plant biology,* 47**,** 127-158.

MARTIN, W. 2010. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365**,** 847-855.

MARTZOUKOU, O., DIALLINAS, G. & AMILLIS, S. 2018. Secretory vesicle polar sorting, endosome recycling and cytoskeleton organization require the AP-1 complex in Aspergillus nidulans. *Genetics,* 209**,** 1121-1138.

MARX, V. 2013. Targeted proteomics. *Nature methods,* 10**,** 19.

MBENGUE, M., BOURDAIS, G., GERVASI, F., BECK, M., ZHOU, J., SPALLEK, T., BARTELS, S., BOLLER, T., UEDA, T. & KUHN, H. 2016. Clathrin-dependent endocytosis is required for immunity mediated by pattern recognition receptor kinases. *Proceedings of the National Academy of Sciences,* 113**,** 11034-11039.

MCDONALD, W. H. & YATES 3RD, J. R. 2003. Shotgun proteomics: integrating technologies to answer biological questions. *Current opinion in molecular therapeutics,* 5**,** 302-309.

MCLAFFERTY, F. W. 1981. Tandem mass spectrometry. *Science,* 214**,** 280-287.

MCLAFFERTY, F. W., BREUKER, K., JIN, M., HAN, X., INFUSINI, G., JIANG, H., KONG, X. & BEGLEY, T. P. 2007. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *The FEBS journal,* 274**,** 6256-6268.

MEIER, I. 2009. Functional Organization of the Plant Nucleus. *In:* MEIER, I. (ed.) *Functional Organization of the Plant Nucleus.* Berlin, Heidelberg: Springer Berlin Heidelberg.

MEIER, I., RICHARDS, E. J. & EVANS, D. E. 2017. Cell biology of the plant nucleus. *Annual review of plant biology,* 68**,** 139-172.

MEINDL, T., BOLLER, T. & FELIX, G. 2000. The bacterial elicitor flagellin activates its receptor in tomato cells according to the address–message concept. *The Plant Cell,* 12**,** 1783-1794.

MERKLE, T. 2011. Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant cell reports,* 30**,** 153-176.

MICHAELSON, D., ABIDI, W., GUARDAVACCARO, D., ZHOU, M., AHEARN, I., PAGANO, M. & PHILIPS, M. R. 2008. Rac1 accumulates in the nucleus during the G2 phase of the cell cycle and promotes cell division. *J Cell Biol,* 181**,** 485-96.

MILNER, E., BARNEA, E., BEER, I. & ADMON, A. 2006. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Molecular & Cellular Proteomics,* 5**,** 357-365.

MISHRA, S., MURPHY, L. C. & MURPHY, L. J. 2006. The Prohibitins: emerging roles in diverse functions. *Journal of cellular and molecular medicine,* 10**,** 353-363.

MIYA, A., ALBERT, P., SHINYA, T., DESAKI, Y., ICHIMURA, K., SHIRASU, K., NARUSAKA, Y., KAWAKAMI, N., KAKU, H. & SHIBUYA, N. 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proceedings of the National Academy of Sciences,* 104**,** 19613-19618.

MOHR, D., FREY, S., FISCHER, T., GUTTLER, T. & GORLICH, D. 2009. Characterisation of the passive permeability barrier of nuclear pore complexes. *EMBO J,* 28**,** 2541-53.

MOSAMMAPARAST, N. & PEMBERTON, L. F. 2004. Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends in cell biology,* 14**,** 547-556.

MOTION, G. B., AMARO, T. M., KULAGINA, N. & HUITEMA, E. 2015. Nuclear processes associated with plant immunity and pathogen susceptibility. *Briefings in functional genomics,* 14**,** 243-252.

NAGARAJ, N., WISNIEWSKI, J. R., GEIGER, T., COX, J., KIRCHER, M., KELSO, J., PÄÄBO, S. & MANN, M. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology,* 7**,** 548.

NARULA, K., CHOUDHARY, P., GHOSH, S., ELAGAMEY, E., CHAKRABORTY, N. & CHAKRABORTY, S. 2019. Comparative Nuclear Proteomics Analysis Provides Insight into the Mechanism of Signaling and Immune Response to Blast Disease Caused by Magnaporthe oryzae in Rice. *Proteomics,* 19**,** e1800188.

NARULA, K., DATTA, A., CHAKRABORTY, N. & CHAKRABORTY, S. 2013. Comparative analyses of nuclear proteome: extending its function. *Frontiers in Plant Science,* 4**,** 100.

NEILSON, K. A., ALI, N. A., MURALIDHARAN, S., MIRZAEI, M., MARIANI, M., ASSADOURIAN, G., LEE, A., VAN SLUYTER, S. C. & HAYNES, P. A. 2011. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics,* 11**,** 535-553.

NI, J.-Q., LIU, L.-P., HESS, D., RIETDORF, J. & SUN, F.-L. 2006. Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. *Genes & development,* 20**,** 1959-1973.

NISHAD, R., AHMED, T., RAHMAN, V. J. & KAREEM, A. 2020. Modulation of plant defense system in response to microbial interactions. *Frontiers in Microbiology,* 11**,** 1298.

NOMURA, H., KOMORI, T., KOBORI, M., NAKAHIRA, Y. & SHIINA, T. 2008. Evidence for chloroplast control of external Ca2+-induced cytosolic Ca2+ transients and stomatal closure. *The Plant Journal,* 53**,** 988-998.

NOMURA, H., KOMORI, T., UEMURA, S., KANDA, Y., SHIMOTANI, K., NAKAI, K., FURUICHI, T., TAKEBAYASHI, K., SUGIMOTO, T. & SANO, S. 2012. Chloroplast-mediated activation of plant immune signalling in Arabidopsis. *Nature communications,* 3**,** 1-11.

O'FARRELL, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *Journal of biological chemistry,* 250**,** 4007-4021.

OHMORI, K., ENDO, Y., YOSHIDA, Y., OHATA, H., TAYA, Y. & ENARI, M. 2008. Monomeric but not trimeric clathrin heavy chain regulates p53-mediated transcription. *Oncogene,* 27**,** 2215-2227.

OHTA, M., SUGITA, M. & SUGIURA, M. 1995. Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in Arabidopsis thaliana: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. *Plant molecular biology,* 27**,** 529-539.

OKSVOLD, M. P., PEDERSEN, N. M., FORFANG, L. & SMELAND, E. B. 2012. Effect of cycloheximide on epidermal growth factor receptor trafficking and signaling. *FEBS letters,* 586**,** 3575-3581.

OLD, W. M., MEYER-ARENDT, K., AVELINE-WOLF, L., PIERCE, K. G., MENDOZA, A., SEVINSKY, J. R., RESING, K. A. & AHN, N. G. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics* S. *Molecular & cellular proteomics,* 4**,** 1487-1502.

ONG, S.-E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics,* 1**,** 376-386.

ONG, S.-E. & MANN, M. 2005. Mass spectrometry–based proteomics turns quantitative. *Nature chemical biology,* 1**,** 252-262.

OOME, S. & VAN DEN ACKERVEKEN, G. 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. *Molecular plant-microbe interactions,* 27**,** 1081-1094.

ORTIZ-MOREA, F. A., SAVATIN, D. V., DEJONGHE, W., KUMAR, R., LUO, Y., ADAMOWSKI, M., VAN DEN BEGIN, J., DRESSANO, K., PEREIRA DE OLIVEIRA, G. & ZHAO, X. 2016. Danger-associated peptide signaling in Arabidopsis requires clathrin. *Proceedings of the National Academy of Sciences,* 113**,** 11028-11033.

PADMANABHAN, M. S. & DINESH-KUMAR, S. 2010. All hands on deck—the role of chloroplasts, endoplasmic reticulum, and the nucleus in driving plant innate immunity. *Molecular Plant-Microbe Interactions,* 23**,** 1368-1380.

PALM, D., SIMM, S., DARM, K., WEIS, B. L., RUPRECHT, M., SCHLEIFF, E. & SCHARF, C. 2016. Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in Arabidopsis thaliana. *RNA Biol,* 13**,** 441-54.

PARK, C.-J., CADDELL, D. F. & RONALD, P. C. 2012. Protein phosphorylation in plant immunity: insights into the regulation of pattern recognition receptor-mediated signaling. *Frontiers in plant science,* 3**,** 177.

PASCAL, S. & SCALLA, R. 1999. Purification and characterization of a safener-induced glutathione S-transferase from wheat (Triticum aestivum). *Physiologia Plantarum,* 106**,** 17-27.

PEETERS, N. & SMALL, I. 2001. Dual targeting to mitochondria and chloroplasts. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1541**,** 54-63.

PENDLE, A. F., CLARK, G. P., BOON, R., LEWANDOWSKA, D., LAM, Y. W., ANDERSEN, J., MANN, M., LAMOND, A. I., BROWN, J. W. S. & SHAW, P. J. 2005. Proteomic Analysis of the Arabidopsis Nucleolus Suggests Novel Nucleolar Functions. *Molecular Biology of the Cell,* 16**,** 260-269.

PENG, Y. T., CHEN, P., OUYANG, R. Y. & SONG, L. 2015. Multifaceted role of prohibitin in cell survival and apoptosis. *Apoptosis,* 20**,** 1135-49.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS: An International Journal,* 20**,** 3551-3567.

PETERSON, A. C., RUSSELL, J. D., BAILEY, D. J., WESTPHALL, M. S. & COON, J. J. 2012. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & Cellular Proteomics,* 11**,** 1475-1488.

PETROVSKÁ, B., ŠEBELA, M. & DOLEŽEL, J. 2015. Inside a plant nucleus: discovering the proteins. *Journal of Experimental Botany,* 66**,** 1627-1640.

PIETERSE, C. M., VAN DER DOES, A., ZAMIOUDIS, C., LEON REYES, H. & VAN WEES, S. C. 2012. Hormonal modulation of plant immunity. *Annual review of cell and developmental biology,* 28**,** 489-521.

PITTERI, S. J., CHRISMAN, P. A., HOGAN, J. M. & MCLUCKEY, S. A. 2005. Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2•. *Analytical chemistry,* 77**,** 1831-1839.

QUTOB, D., KEMMERLING, B., BRUNNER, F., KUFNER, I., ENGELHARDT, S., GUST, A. A., LUBERACKI, B., SEITZ, H. U., STAHL, D. & RAUHUT, T. 2006. Phytotoxicity and innate immune responses induced by Nep1-like proteins. *The Plant Cell,* 18**,** 3721-3744.

RAFFAELE, S. & RIVAS, S. 2013. Regulate and be regulated: integration of defense and other signals by the AtMYB30 transcription factor. *Frontiers in plant science,* 4**,** 98.

RAJAMAKI, M. L., SIKORSKAITE-GUDZIUNIENE, S., SARMAH, N., VARJOSALO, M. & VALKONEN, J. P. T. 2020. Nuclear proteome of virus-infected and healthy potato leaves. *BMC Plant Biol,* 20**,** 355.

RANF, S. 2017. Sensing of molecular patterns through cell surface immune receptors. *Current Opinion in Plant Biology,* 38**,** 68-77.

RIBBECK, K. & GÖRLICH, D. 2001. Kinetic analysis of translocation through nuclear pore complexes. *The EMBO journal,* 20**,** 1320-1330.

RIVAS, S. & DESLANDES, L. 2013. Nuclear components and dynamics during plant innate immunity. Frontiers Media SA.

ROBATZEK, S., CHINCHILLA, D. & BOLLER, T. 2006. Ligand-induced endocytosis of the pattern recognition receptor FLS2 in Arabidopsis. *Genes & development,* 20**,** 537-542.

ROESLI, C., NERI, D. & RYBAK, J.-N. 2006. In vivo protein biotinylation and sample preparation for the proteomic identification of organ-and disease-specific antigens accessible from the vasculature. *Nature protocols,* 1**,** 192-199.

ROSS, P. L., HUANG, Y. N., MARCHESE, J. N., WILLIAMSON, B., PARKER, K., HATTAN, S., KHAINOVSKI, N., PILLAI, S., DEY, S. & DANIELS, S. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics,* 3**,** 1154-1169.

ROUX, K. J., KIM, D. I., RAIDA, M. & BURKE, B. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology,* 196**,** 801-810.

SÁEZ-VÁSQUEZ, J. & DELSENY, M. 2019. Ribosome Biogenesis in Plants: From Functional 45S Ribosomal DNA Organization to Ribosome Assembly Factors. *The Plant Cell,* 31**,** 1945.

SAIJO, Y., LOO, E. P. I. & YASUDA, S. 2018. Pattern recognition receptors and signaling in plant–microbe interactions. *The Plant Journal,* 93**,** 592-613.

SAKAMOTO, Y. & TAKAGI, S. 2013. LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana. *Plant Cell Physiol,* 54**,** 622-33.

ŠAMAJOVÁ, O., KOMIS, G. & ŠAMAJ, J. 2013. Emerging topics in the cell biology of mitogen-activated protein kinases. *Trends in Plant Science,* 18**,** 140-148.

SANGER, F. 1959. Chemistry of insulin: determination of the structure of insulin opens the way to greater understanding of life processes. *Science,* 129**,** 1340-1344.

SATTLER, S. E. & FUNNELL-HARRIS, D. L. 2013. Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens? *Frontiers in plant science,* 4**,** 70.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature protocols,* 3**,** 1101-1108.

SCHÖN, M., TÖLLER, A., DIEZEL, C., ROTH, C., WESTPHAL, L., WIERMER, M. & SOMSSICH, I. E. 2013. Analyses of wrky18 wrky40 plants reveal critical roles of SA/EDS1 signaling and indole-glucosinolate biosynthesis for Golovinomyces orontii resistance and a loss-of resistance towards Pseudomonas syringae pv. tomato AvrRPS4. *Molecular Plant-Microbe Interactions,* 26**,** 758-767.

SCHRÖDER, P., SCHEER, C. E., DIEKMANN, F. & STAMPFL, A. 2007. How plants cope with foreign compounds. Translocation of xenobiotic glutathione conjugates in roots of barley (Hordeum vulgare)(9 pp). *Environmental Science and Pollution Research-International,* 14**,** 114-122.

SCHWACKE, R., FISCHER, K., KETELSEN, B., KRUPINSKA, K. & KRAUSE, K. 2007. Comparative survey of plastid and mitochondrial targeting properties of transcription factors in Arabidopsis and rice. *Molecular Genetics and Genomics,* 277**,** 631-646.

SEGUEL, A., JELENSKA, J., HERRERA-VASQUEZ, A., MARR, S. K., JOYCE, M. B., GAGESCH, K. R., SHAKOOR, N., JIANG, S. C., FONSECA, A., WILDERMUTH, M. C., GREENBERG, J. T. & HOLUIGUE, L. 2018. PROHIBITIN3 Forms Complexes with ISOCHORISMATE SYNTHASE1 to Regulate Stress-Induced Salicylic Acid Biosynthesis in Arabidopsis. *Plant Physiol,* 176**,** 2515-2531.

SEO, Y. R., FISHEL, M. L., AMUNDSON, S., KELLEY, M. R. & SMITH, M. L. 2002. Implication of p53 in base excision DNA repair: in vivo evidence. *Oncogene,* 21**,** 731-737.

SHARMA, M., BENNEWITZ, B. & KLÖSGEN, R. B. 2018. Rather rule than exception? How to evaluate the relevance of dual protein targeting to mitochondria and chloroplasts. *Photosynthesis Research,* 138**,** 335-343.

SHERMAN, B. T., HAO, M., QIU, J., JIAO, X., BASELER, M. W., LANE, H. C., IMAMICHI, T. & CHANG, W. 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res,* 10.

SHEVCHENKO, A., JENSEN, O. N., PODTELEJNIKOV, A. V., SAGLIOCCO, F., WILM, M., VORM, O., MORTENSEN, P., SHEVCHENKO, A., BOUCHERIE, H. & MANN, M. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences,* 93**,** 14440-14445.

SHUKLA, A. K. & FUTRELL, J. H. 2000. Tandem mass spectrometry: dissociation of ions by collisional activation. *Journal of mass spectrometry,* 35**,** 1069-1090.

SIKORSKAITE-GUDZIUNIENE, S., HAIMI, P., GELVONAUSKIENE, D. & STANYS, V. 2017. Nuclear proteome analysis of apple cultivar 'Antonovka'accessions in response to apple scab (Venturia inaequalis). *European Journal of Plant Pathology,* 148**,** 771-784.

SOBOČANEC, S., FILIĆ, V., MATOVINA, M., MAJHEN, D., ŠAFRANKO, Ž. M., HADŽIJA, M. P., KRSNIK, Ž., KURILJ, A. G., ŠARIĆ, A. & ABRAMIĆ, M. 2016. Prominent role of exopeptidase DPP III in estrogen-mediated protection against hyperoxia in vivo. *Redox biology,* 8**,** 149-159.

SOLANO DE LA CRUZ, M. T., ADAME-GARCÍA, J., GREGORIO-JORGE, J., JIMÉNEZ-JACINTO, V., VEGA-ALVARADO, L., IGLESIAS-ANDREU, L., ESCOBAR-HERNÁNDEZ, E. E. & LUNA-RODRÍGUEZ, M. 2019. Increase in ribosomal proteins activity: Translational reprogramming in Vanilla planifolia Jacks., against Fusarium infection. bioRxiv.

SONG, G., HSU, P. Y. & WALLEY, J. W. 2018. Assessment and refinement of sample preparation methods for deep and quantitative plant proteome profiling. *Proteomics,* 18**,** 1800220.

SOSA, B. A., ROTHBALLER, A., KUTAY, U. & SCHWARTZ, T. U. 2012. LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins. *Cell,* 149**,** 1035-1047.

SPALLEK, T., BECK, M., BEN KHALED, S., SALOMON, S., BOURDAIS, G., SCHELLMANN, S. & ROBATZEK, S. 2013. ESCRT-I mediates FLS2 endosomal sorting and plant immunity. *PLoS genetics,* 9**,** e1004035.

SPERSCHNEIDER, J., CATANZARITI, A.-M., DEBOER, K., PETRE, B., GARDINER, D. M., SINGH, K. B., DODDS, P. N. & TAYLOR, J. M. 2017. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports,* 7**,** 44598.

STEEN, H. & MANN, M. 2004. The ABC's (and XYZ's) of peptide sequencing. *Nature reviews Molecular cell biology,* 5**,** 699-711.

STĘPIŃSKI, D. 2014. Functional ultrastructure of the plant nucleolus. *Protoplasma,* 251**,** 1285-1306.

STEWART, M. 2007. Molecular mechanism of the nuclear protein import cycle. *Nature reviews Molecular cell biology,* 8**,** 195-208.

SU, J., YANG, L., ZHU, Q., WU, H., HE, Y., LIU, Y., XU, J., JIANG, D. & ZHANG, S. 2018. Active photosynthetic inhibition mediated by MPK3/MPK6 is critical to effector-triggered immunity. *PLoS Biology,* 16**,** e2004122.

SUGANUMA, T., PATTENDEN, S. G. & WORKMAN, J. L. 2008. Diverse functions of WD40 repeat proteins in histone recognition. *Genes Dev,* 22**,** 1265-8.

SUNDERLAND, P. A., WEST, C. E., WATERWORTH, W. M. & BRAY, C. M. 2006. An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in Arabidopsis thaliana. *The Plant Journal,* 47**,** 356-367.

SUTHERLAND, H. G., MUMFORD, G. K., NEWTON, K., FORD, L. V., FARRALL, R., DELLAIRE, G., CÁCERES, J. F. & BICKMORE, W. A. 2001. Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Human Molecular Genetics,* 10**,** 1995-2011.

SYKA, J. E., COON, J. J., SCHROEDER, M. J., SHABANOWITZ, J. & HUNT, D. F. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences,* 101**,** 9528-9533.

SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H. & BORK, P. 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research,* 47**,** D607-D613.

TAKAHASHI, A., KAWASAKI, T., HENMI, K., SHII, K., KODAMA, O., SATOH, H. & SHIMAMOTO, K. 1999. Lesion mimic mutants of rice with alterations in early signaling events of defense. *The Plant journal : for cell and molecular biology,* 17 5**,** 535-45.

TAKAHASHI, A., KAWASAKI, T., WONG, H. L., SUHARSONO, U., HIRANO, H. & SHIMAMOTO, K. 2003. Hyperphosphorylation of a Mitochondrial Protein, Prohibitin, Is Induced by Calyculin A in a Rice Lesion-Mimic Mutant cdr1. *Plant Physiology,* 132**,** 1861.

TAMURA, K., FUKAO, Y., IWAMOTO, M., HARAGUCHI, T. & HARA-NISHIMURA, I. 2010. Identification and characterization of nuclear pore complex components in Arabidopsis thaliana. *The Plant Cell,* 22**,** 4084-4097.

TAMURA, K. & HARA-NISHIMURA, I. 2013. The molecular architecture of the plant nuclear pore complex. *Journal of experimental botany,* 64**,** 823-832.

TAMURA, K. & HARA-NISHIMURA, I. 2014. Functional insights of nucleocytoplasmic transport in plants. *Front Plant Sci,* 5**,** 118.

TANG, Y., HUANG, A. & GU, Y. 2020. Global profiling of plant nuclear membrane proteome in Arabidopsis. *Nat Plants,* 6**,** 838-847.

TANZ, S. K., CASTLEDEN, I., SMALL, I. D. & MILLAR, A. H. 2013. Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Frontiers in Plant Science,* 4**,** 214.

TCHÓRZEWSKI, M., BOLDYREFF, B. & GRANKOWSKI, N. 1999. Extraribosomal function of the acidic ribosomal P1-protein YP1alpha from Saccharomyces cerevisiae. *Acta biochimica Polonica,* 46**,** 901-910.

THOMPSON, A., SCHÄFER, J., KUHN, K., KIENLE, S., SCHWARZ, J., SCHMIDT, G., NEUMANN, T. & HAMON, C. 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry,* 75**,** 1895-1904.

THOMSON, E., FERREIRA-CERCA, S. & HURT, E. 2013. Eukaryotic ribosome biogenesis at a glance. *Journal of Cell Science,* 126**,** 4815.

THUAUD, F., RIBEIRO, N., NEBIGIL, CANAN G. & DÉSAUBRY, L. 2013. Prohibitin Ligands in Cell Death and Survival: Mode of Action and Therapeutic Potential. *Chemistry & Biology,* 20**,** 316-331.

THUL, P. J., ÅKESSON, L., WIKING, M., MAHDESSIAN, D., GELADAKI, A., AIT BLAL, H., ALM, T., ASPLUND, A., BJÖRK, L. & BRECKELS, L. M. 2017. A subcellular map of the human proteome. *Science,* 356**,** eaal3321.

TIAN, F., YANG, D.-C., MENG, Y.-Q., JIN, J. & GAO, G. 2019. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research,* 48**,** D1104-D1113.

TIMNEY, B. L., RAVEH, B., MIRONSKA, R., TRIVEDI, J. M., KIM, S. J., RUSSEL, D., WENTE, S. R., SALI, A. & ROUT, M. P. 2016. Simple rules for passive diffusion through the nuclear pore complex. *J Cell Biol,* 215**,** 57-76.

TORTOSA, M., CARTEA, M. E., VELASCO, P., SOENGAS, P. & RODRIGUEZ, V. M. 2019. Calcium-signaling proteins mediate the plant transcriptomic response during a well-established Xanthomonas campestris pv. campestris infection. *Horticulture research,* 6.

TU, W.-Y., HUANG, Y.-C., LIU, L., CHANG, L.-H. & TAM, M. 2011. Rpl12p affects the transcription of the PHO pathway high affinity inorganic phosphate transporters and repressible phosphatases. *Yeast,* 28.

TUROWSKI, T. W. & TOLLERVEY, D. 2015. Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley Interdiscip Rev RNA,* 6**,** 129-39.

TYANOVA, S., TEMU, T., SINITCYN, P., CARLSON, A., HEIN, M. Y., GEIGER, T., MANN, M. & COX, J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods,* 13**,** 731-740.

TYERS, M. & MANN, M. 2003. From genomics to proteomics. *Nature,* 422**,** 193-197.

UNDERWOOD, W. 2012. The plant cell wall: a dynamic barrier against pathogen invasion. *Frontiers in plant science,* 3**,** 85.

VADOVIC, P., SAMAJOVA, O., TAKAC, T., NOVAK, D., ZAPLETALOVA, V., COLCOMBET, J. & SAMAJ, J. 2019. Biochemical and Genetic Interactions of Phospholipase D Alpha 1 and Mitogen-Activated Protein Kinase 3 Affect Arabidopsis Stress Response. *Front Plant Sci,* 10**,** 275.

VAINONEN, J. P., SAKURAGI, Y., STAEL, S., TIKKANEN, M., ALLAHVERDIYEVA, Y., PAAKKARINEN, V., ARO, E., SUORSA, M., SCHELLER, H. V. & VENER, A. V. 2008. Light regulation of CaS, a novel phosphoprotein in the thylakoid membrane of Arabidopsis thaliana. *The FEBS journal,* 275**,** 1767-1777.

VAN AKEN, O., PECENKOVÁ, T., VAN DE COTTE, B., DE RYCKE, R., EECKHOUT, D., FROMM, H., DE JAEGER, G., WITTERS, E., BEEMSTER, G. T. S., INZÉ, D. & VAN BREUSEGEM, F. 2007. Mitochondrial type-I prohibitins of Arabidopsis thaliana are required for supporting proficient meristem development. *The Plant journal : for cell and molecular biology,* 52**,** 850-864.

VAN AKEN, O., WHELAN, J. & VAN BREUSEGEM, F. 2010. Prohibitins: mitochondrial partners in development and stress response. *Trends in Plant Science,* 15**,** 275-282.

VAN AKEN, O., ZHANG, B., LAW, S., NARSAI, R. & WHELAN, J. 2013. AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. *Plant physiology,* 162**,** 254-271.

VENABLE, J. D., DONG, M.-Q., WOHLSCHLEGEL, J., DILLIN, A. & YATES, J. R. 2004. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods,* 1**,** 39-45.

VINAIPHAT, A., LOW, J. K., YEOH, K. W., CHNG, W. J. & SZE, S. K. 2021. Application of advanced mass spectrometry-based proteomics to study hypoxia driven cancer progression. *Frontiers in Oncology,* 11**,** 98.

WAN, W.-L., ZHANG, L., PRUITT, R., ZAIDEM, M., BRUGMAN, R., MA, X., KROL, E., PERRAKI, A., KILIAN, J., GROSSMANN, G., STAHL, M., SHAN, L., ZIPFEL, C., VAN KAN, J. A. L., HEDRICH, R., WEIGEL, D., GUST, A. A. & NÜRNBERGER, T. 2019. Comparing Arabidopsis receptor kinase and receptor protein-mediated immune signaling reveals BIK1-dependent differences. *The New phytologist,* 221**,** 2080-2095.

WANG, J., HU, M., WANG, J., QI, J., HAN, Z., WANG, G., QI, Y., WANG, H.-W., ZHOU, J.-M. & CHAI, J. 2019. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science,* 364**,** eaav5870.

WANG, R. & BRATTAIN, M. G. 2007. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Lett,* 581**,** 3164-70.

WANG, X. 2005. Regulatory functions of phospholipase D and phosphatidic acid in plant growth, development, and stress responses. *Plant physiology,* 139**,** 566-573.

WAR, A. R., PAULRAJ, M. G., AHMAD, T., BUHROO, A. A., HUSSAIN, B., IGNACIMUTHU, S. & SHARMA, H. C. 2012. Mechanisms of plant defense against insect herbivores. *Plant signaling & behavior,* 7**,** 1306-1320.

WATKINS, N. J. & BOHNSACK, M. T. 2012. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA,* 3**,** 397-414.

WEINL, S., HELD, K., SCHLÜCKING, K., STEINHORST, L., KUHLGERT, S., HIPPLER, M. & KUDLA, J. 2008. A plastid protein crucial for Ca2+-regulated stomatal responses. *New Phytologist,* 179**,** 675-686.

WEIS, K. 2003. Regulating Access to the Genome: Nucleocytoplasmic Transport throughout the Cell Cycle. *Cell,* 112**,** 441-451.

WIATROWSKI, H. A. & CARLSON, M. 2003. Yap1 accumulates in the nucleus in response to carbon stress in Saccharomyces cerevisiae. *Eukaryot Cell,* 2**,** 19-26.

WIESE, S., REIDEGELD, K. A., MEYER, H. E. & WARSCHEID, B. 2007. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics,* 7**,** 340-350.

WIŚNIEWSKI, J. R., HEIN, M. Y., COX, J. & MANN, M. 2014. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Molecular & cellular proteomics,* 13**,** 3497-3506.

WIŚNIEWSKI, J. R., ZOUGMAN, A., NAGARAJ, N. & MANN, M. 2009. Universal sample preparation method for proteome analysis. *Nature methods,* 6**,** 359-362.

WOOLFORD, J. L., JR. & BASERGA, S. J. 2013. Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics,* 195**,** 643-681.

WU, Y., ZHANG, D., CHU, J. Y., BOYLE, P., WANG, Y., BRINDLE, I. D., DE LUCA, V. & DESPRÉS, C. 2012. The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell reports,* 1**,** 639-647.

YAGER, T. D., NICKERSON, D. A. & HOOD, L. E. 1991. The human genome project: creating an infrastructure for biology and medicine. *Trends in Biochemical Sciences,* 16**,** 454.

YAN, X., WANG, Y., XU, M., DAHHAN, D. A., LIU, C., ZHANG, Y., LIN, J., BEDNAREK, S. Y. & PAN, J. 2021. Cross-talk between clathrin-dependent post-Golgi trafficking and clathrin-mediated endocytosis in Arabidopsis root cells. *The Plant Cell,* 33**,** 3057-3075.

YANG, Y., WANG, W., CHU, Z., ZHU, J. K. & ZHANG, H. 2017. Roles of Nuclear Pores and Nucleo-cytoplasmic Trafficking in Plant Stress Responses. *Front Plant Sci,* 8**,** 574.

YEATS, T. H. & ROSE, J. K. 2013. The formation and function of plant cuticles. *Plant physiology,* 163**,** 5-20.

YIN, X. & KOMATSU, S. 2016. Plant nuclear proteomics for unraveling physiological function. *New biotechnology,* 33**,** 644-654.

YOCGO, R. E., GEZA, E., CHIMUSA, E. R. & MAZANDU, G. K. 2017. A post-gene silencing bioinformatics protocol for plant-defence gene validation and underlying process identification: case study of the Arabidopsis thaliana NPR1. *BMC plant biology,* 17**,** 1-18.

YOUNG, S. A., WANG, X. & LEACH, J. E. 1996. Changes in the Plasma Membrane Distribution of Rice Phospholipase D during Resistant Interactions with Xanthomonas oryzae pv oryzae. *The Plant Cell,* 8**,** 1079.

YU, L.-R., STEWART, N. A. & VEENSTRA, T. D. 2010. Chapter 8 - Proteomics: The Deciphering of the Functional Genome. *In:* GINSBURG, G. S. & WILLARD, H. F. (eds.) *Essentials of Genomic and Personalized Medicine.* San Diego: Academic Press.

YU, Y. 2019. Prohibitin Shuttles Between Mitochondria and the Nucleus to Control Genome Stability During the Cell Cycle. *Plant Physiol,* 179**,** 1435-1436.

ZHANG, J., LI, W., XIANG, T., LIU, Z., LALUK, K., DING, X., ZOU, Y., GAO, M., ZHANG, X. & CHEN, S. 2010. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. *Cell host & microbe,* 7**,** 290-301.

ZHANG, Y., SONG, G., LAL, N. K., NAGALAKSHMI, U., LI, Y., ZHENG, W., HUANG, P.-J., BRANON, T. C., TING, A. Y. & WALLEY, J. W. 2019. TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. *Nature communications,* 10**,** 1-17.

ZHAO, J., DEVAIAH, S. P., WANG, C., LI, M., WELTI, R. & WANG, X. 2013. Arabidopsis phospholipase Dβ1 modulates defense responses to bacterial and fungal pathogens. *The New phytologist,* 199**,** 228-240.

ZHENG, H.-C. & JIANG, H.-M. 2022. Shuttling of cellular proteins between the plasma membrane and nucleus. *Molecular Medicine Reports,* 25**,** 1-12.

ZHOU, J.-M. & ZHANG, Y. 2020. Plant immunity: danger perception and signaling. *Cell,* 181**,** 978-989.

ZHOU, X., GRAUMANN, K., EVANS, D. E. & MEIER, I. 2012. Novel plant SUN–KASH bridges are involved in RanGAP anchoring and nuclear shape determination. *Journal of Cell Biology,* 196**,** 203-211.

ZHOU, X., LIAO, W. J., LIAO, J. M., LIAO, P. & LU, H. 2015. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol,* 7**,** 92-104.

ZNAIDI, S., PELLETIER, B., MUKAI, Y. & LABBE, S. 2004. The Schizosaccharomyces pombe corepressor Tup11 interacts with the iron-responsive transcription factor Fep1. *J Biol Chem,* 279**,** 9462-74.

ZUBAREV, R. 2006. Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. *Expert review of proteomics,* 3**,** 251-261.

ZUBAREV, R. A., KELLEHER, N. L. & MCLAFFERTY, F. W. 1998. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society,* 120**,** 3265-3266.

8. Appendix

6.1 Appendix 1

All supplementary tables and files in appendix 1 are found on the CD submitted with this thesis. These data is so big and cannot be added to the word file.

Supplementary file 1. (Tables 1-10)

Supplementary file 2. (Tables 1-12)

6.2 Appendix 2

Supplementary table 1. Proteins annotated to the transcription process were classified into families. UP_KEYWORDS category was used in the David bioinformatics resources annotation. Control and elicited conditions were used to generate this list.

Transcription Factors families			
	Number of proteins		
AP2	2	AT5G13330	related to AP2 6l(Rap2.6L)
		AT3G18990	AP2/B3-like transcriptional factor family protein(VRN1)
BES1	2	AT1G19350	Brassinosteroid signaling positive regulator (BZR1) family protein(BES1)
		AT1G69010	BES1-interacting Myc-like protein 2(BIM2)
BSD	1	AT1G55750	BSD domain (BTF2-like transcription factors, Synapse-associated proteins and DOS2-like proteins)(AT1G55750)
bZIP	13	AT1G45249	abscisic acid responsive elements-binding factor 2(ABF2)
		AT4G34000	abscisic acid responsive elements-binding factor 3(ABF3)
		AT5G06950	bZIP transcription factor family protein(AHBP-1B)
		AT2G35530	basic region/leucine zipper transcription factor 16(bZIP16)
		AT1G32150	basic region/leucine zipper transcription factor 68(bZIP68)
		AT2G31370	Basic-leucine zipper (bZIP) transcription factor family protein(AT2G31370)
		AT2G36270	Basic-leucine zipper (bZIP) transcription factor family protein(ABI5)
		AT5G11260	Basic-leucine zipper (bZIP) transcription factor family protein(HY5)
		AT3G56850	ABA-responsive element binding protein 3(AREB3)
		AT3G19290	ABRE binding factor 4(ABF4)
		AT4G36730	G-box binding factor 1(GBF1)
		AT2G46270	G-box binding factor 3(GBF3)
		AT5G06960	OCS-element binding factor 5(OBF5)
C2H2	4	AT5G04240	Zinc finger (C2H2 type) family protein / transcription factor jumonji (jmj) family protein(ELF6)
		AT1G14580	C2H2-like zinc finger protein(AT1G14580)

		AT1G55110	indeterminate(ID)-domain 7(IDD7)
		AT3G48430	relative of early flowering 6(REF6)
E2F	2	AT5G22220	E2F transcription factor 1(E2F1)
		AT5G03415	Transcription factor DP(DPB)
SAP	1	AT2G45640	SIN3 associated polypeptide P18(SAP18)
TCP	2	AT1G72010	TCP family transcription factor(AT1G72010)
		AT1G58100	TCP family transcription factor(TCP8)
ARF	4	AT5G62000	auxin response factor 2(ARF2)
		AT1G30330	auxin response factor 6(ARF6)
		AT5G37020	auxin response factor 8(ARF8)
		AT5G20730	Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-like protein(NPH4)
WRKY	8	AT5G49520	WRKY DNA-binding protein 48(WRKY48)
		AT4G12020	protein kinase family protein(WRKY19)
		AT2G38470	WRKY DNA-binding protein 33(WRKY33)
		AT1G13960	WRKY DNA-binding protein 4(WRKY4)
		AT1G69310	WRKY DNA-binding protein 57(WRKY57)
		AT1G62300	WRKY family transcription factor(WRKY6)
		AT5G45260	Disease resistance protein (TIR-NBS-LRR class)(RRS1)
		AT2G04880	zinc-dependent activator protein-1(ZAP1)
Alfin-like	4	AT5G05610	alfin-like 1(AL1)
		AT3G42790	alfin-like 3(AL3)
		AT2G02470	alfin-like 6(AL6)
		AT1G14510	alfin-like 7(AL7)
bHLH	7	AT1G05805	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G05805)
		AT4G28790	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT4G28790)
		AT1G35460	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH1)
		AT1G51140	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH3)
		AT2G42280	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH4)
		AT4G36060	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(bHLH11)
		AT5G01310	APRATAXIN-like protein(APTX)
BPC	1	AT5G42520	basic pentacysteine 6(BPC6)
global	7	AT4G10710	global transcription factor C(SPT16)
		AT4G08350	global transcription factor group A2(GTA2)
		AT1G65440	global transcription factor group B1(GTB1)
		AT5G10550	global transcription factor group E2(GTE2)
		AT1G73150	global transcription factor group E3(GTE3)
		AT1G06230	global transcription factor group E4(GTE4)
		AT5G65630	global transcription factor group E7(GTE7)
HSF	2	AT4G11660	winged-helix DNA-binding transcription factor family protein(AT-HSFB2B)
		AT3G22830	heat shock transcription factor A6B(HSFA6B)

MYB	4	AT5G47390	myb-like transcription factor family protein(AT5G47390)
		AT2G47210	myb-like transcription factor family protein(AT2G47210)
		AT5G61620	myb-like transcription factor family protein(AT5G61620)
		AT1G09770	cell division cycle 5(CDC5)
MYB related	5	AT4G34430	DNA-binding family protein(CHB3)
		AT1G72740	Homeodomain-like/winged-helix DNA-binding family protein(AT1G72740)
		AT2G47620	SWITCH/sucrose nonfermenting 3A(SWI3A)
		AT1G21700	SWITCH/sucrose nonfermenting 3C(SWI3C)
		AT1G49950	telomere repeat binding factor 1(TRB1)
GRAS	1	AT1G55580	GRAS family transcription factor(LAS)
		AT1G50420	scarecrow-like 3(SCL3)
NAC	5	AT3G10480	NAC domain containing protein 50(NAC050)
		AT3G10490	NAC domain containing protein 52(NAC052)
		AT1G73230	Nascent polypeptide-associated complex NAC(AT1G73230)
		AT4G28530	NAC domain containing protein 74(NAC074)
		AT2G27300	NTM1-like 8(NTL8)
Trihelix	8	AT1G54060	6B-interacting protein 1-like 1(ASIL1)
		AT2G33550	Homeodomain-like superfamily protein(AT2G33550)
		AT2G38250	Homeodomain-like superfamily protein(AT2G38250)
		AT3G04450	Homeodomain-like superfamily protein(AT3G04450)
		AT3G19070	Homeodomain-like superfamily protein(AT3G19070)
		AT5G01380	Homeodomain-like superfamily protein(AT5G01380)
		AT5G29000	Homeodomain-like superfamily protein(PHL1)
		AT3G14180	sequence-specific DNA binding transcription factor(ASIL2)
CAMTA		AT5G64220	Calmodulin-binding transcription activator protein with CG-1 and Ankyrin domain(AT5G64220)
GeBP	5	AT1G61730	DNA-binding storekeeper protein-related transcriptional regulator(AT1G61730)
		AT3G04930	DNA-binding storekeeper protein-related transcriptional regulator(AT3G04930)
		AT4G00390	DNA-binding storekeeper protein-related transcriptional regulator(AT4G00390)
		AT5G28040	DNA-binding storekeeper protein-related transcriptional regulator(AT5G28040)
		AT1G11510	DNA-binding storekeeper protein-related transcriptional regulator(AT1G11510)
HB	6	AT5G44180	Homeodomain-like transcriptional regulator(RLT2)
		AT5G17810	WUSCHEL related homeobox 12(WOX12)
		AT1G27050	homeobox leucine zipper protein(AT1G27050)
		AT1G28420	homeobox-1(HB-1)
		AT3G61150	homeodomain GLABROUS 1(HDG1)
		AT5G66700	homeobox 53(HB53)
AP2-EREBP	1	AT1G51190	Integrase-type DNA-binding superfamily protein(PLT2)
orphans	3	AT5G15020	SIN3-like 2(SNL2)
		AT1G24190	SIN3-like 3(SNL3)
		AT1G70060	SIN3-like 4(SNL4)

C2C2-Dof	1	AT5G39660	cycling DOF factor 2(CDF2)
TAZ	1	AT1G79000	histone acetyltransferase of the CBP family 1(HAC1)
CCAAT	7	AT5G12840	nuclear factor Y, subunit A1(NF-YA1)
		AT2G38880	nuclear factor Y, subunit B1(NF-YB1)
		AT3G53340	nuclear factor Y, subunit B10(NF-YB10)
		AT5G23090	nuclear factor Y, subunit B13(NF-YB13)
		AT1G08970	nuclear factor Y, subunit C9(NF-YC9)
		AT5G47640	nuclear factor Y, subunit B2(NF-YB2)
		AT1G56170	nuclear factor Y, subunit C2(NF-YC2)
ABI3VP1	1	AT1G49480	related to vernalization1 1(RTV1)
Coactivator p15	2	AT5G09250	ssDNA-binding transcriptional regulator(KIWI)
		AT4G10920	transcriptional coactivator p15 (PC4) family protein (KELP)(KELP)
SNF2	1	AT2G46020	transcription regulatory protein SNF2(BRM)
Transcriptional regulators families	Number of proteins		
ARID	3	AT1G20910	ARID/BRIGHT DNA-binding domain-containing protein(AT1G20910)
		AT1G76510	ARID/BRIGHT DNA-binding domain-containing protein(AT1G76510)
		AT2G17410	ARID/BRIGHT DNA-binding domain-containing protein(AT2G17410)
LUG	2	AT2G32700	LEUNIG-like protein(LUH)
		AT4G32551	transcriptional corepressor LEUNIG(LUG)
SWI/SNF-SWI3	1	AT4G16310	LSD1-like 3(LDL3)
PHD	3	AT4G22140	PHD finger family protein / bromo-adjacent homology (BAH) domain-containing protein(EBS)
		AT4G39100	PHD finger family protein / bromo-adjacent homology (BAH) domain-containing protein(SHL1)
		AT4G12620	origin of replication complex 1B(ORC1B)
SWI/SNF-BAF60b	1	AT5G14170	SWIB/MDM2 domain superfamily protein(CHC1)
HMG	1	AT3G28730	high mobility group(HMG)
GNAT	1	AT3G54610	histone acetyltransferase of the GNAT family 1(HAG1)
Jumonji	1	AT4G20400	JUMONJI 14(JMJ14)
MBF1	1	AT2G42680	multiprotein bridging factor 1A(MBF1A)
Others	Number of proteins		
	131	AT1G63470	AT hook motif DNA-binding family protein(AT1G63470)
		AT1G63480	AT hook motif DNA-binding family protein(AT1G63480)
		AT2G33620	AT hook motif DNA-binding family protein(AT2G33620)
		AT3G04590	AT hook motif DNA-binding family protein(AT3G04590)
		AT3G61310	AT hook motif DNA-binding family protein(AT3G61310)
		AT4G00200	AT hook motif DNA-binding family protein(AT4G00200)
		AT4G17950	AT hook motif DNA-binding family protein(AT4G17950)
		AT4G22770	AT hook motif DNA-binding family protein(AT4G22770)
		AT4G25320	AT hook motif DNA-binding family protein(AT4G25320)

		AT5G46640	AT hook motif DNA-binding family protein(AT5G46640)
		AT5G51590	AT hook motif DNA-binding family protein(AT5G51590)
		AT5G62260	AT hook motif DNA-binding family protein(AT5G62260)
		AT4G12080	AT-hook motif nuclear-localized protein 1(AHL1)
		AT3G04570	AT-hook motif nuclear-localized protein 19(AHL19)
		AT2G45430	AT-hook motif nuclear-localized protein 22(AHL22)
		AT2G32940	Argonaute family protein(AGO6)
		AT5G15160	BANQUO 2(BNQ2)
		AT3G42170	BED zinc finger and hAT dimerization domain-containing protein DAYSLEEPER(DAYSLEEPER)
		AT4G21670	C-terminal domain phosphatase-like 1(CPL1)
		AT1G18950	DDT domain superfamily(AT1G18950)
		AT3G13940	DNA binding / DNA-directed RNA polymerase(AT3G13940)
		AT1G60850	DNA-directed RNA polymerase family protein(ATRPAC42)
		AT4G21710	DNA-directed RNA polymerase family protein(NRPB2)
		AT2G15430	DNA-directed RNA polymerase family protein(NRPB3)
		AT5G05130	DNA/RNA helicase protein(AT5G05130)
		AT3G58560	DNAse I-like superfamily protein(AT3G58560)
		AT3G58580	DNAse I-like superfamily protein(AT3G58580)
		AT5G11350	DNAse I-like superfamily protein(AT5G11350)
		AT3G22320	Eukaryotic rpb5 RNA polymerase subunit family protein(NRPB5)
		AT5G08550	GC-rich sequence DNA-binding factor-like protein(ILP1)
		AT2G16485	GW repeat- and PHD finger-containing protein NERD(NERD)
		AT5G42020	Heat shock protein 70 (Hsp 70) family protein(BIP2)
		AT5G02490	Heat shock protein 70 (Hsp 70) family protein(Hsp70-2)
		AT5G13680	IKI3 family protein(ABO1)
		AT1G07090	LIGHT-DEPENDENT SHORT HYPOCOTYLS-like protein (DUF640)(LSH6)
		AT4G02060	Minichromosome maintenance (MCM2/3/5) family protein(PRL)
		AT1G10170	NF-X-like 1(NFXL1)
		AT5G67630	P-loop containing nucleoside triphosphate hydrolases superfamily protein(AT5G67630)
		AT5G22330	P-loop containing nucleoside triphosphate hydrolases superfamily protein(RIN1)
		AT2G28290	P-loop containing nucleoside triphosphate hydrolases superfamily protein(SYD)
		AT3G22590	PLANT HOMOLOGOUS TO PARAFIBROMIN(PHP)
		AT5G60960	Pentatricopeptide repeat (PPR) superfamily protein(PNM1)
		AT1G76500	Putative AT-hook DNA-binding family protein(SOB3)
		AT5G44280	RING 1A(RING1A)
		AT2G43410	RNA binding protein(FPA)
		AT5G13010	RNA helicase family protein(EMB3011)
		AT3G46960	RNA helicase, ATP-dependent, SK12/DOB1 protein(AT3G46960)
		AT4G35800	RNA polymerase II large subunit(NRPB1)
		AT1G55325	RNA polymerase II transcription mediator(GCT)

		AT5G09920	RNA polymerase II, Rpb4, core protein(NRPB4)
		AT2G04630	RNA polymerase Rpb6(NRPB6B)
		AT3G16980	RNA polymerases M/15 Kd subunit(NRPB9A)
		AT3G57660	nuclear RNA polymerase A1(NRPA1)
		AT1G29940	nuclear RNA polymerase A2(NRPA2)
		AT5G60040	nuclear RNA polymerase C1(NRPC1)
		AT3G04610	RNA-binding KH domain-containing protein(FLK)
		AT4G26000	RNA-binding KH domain-containing protein(PEP)
		AT5G46250	RNA-binding protein(LARP6a)
		AT2G43970	RNA-binding protein(LARP6b)
		AT2G29540	RNApolymerase 14 kDa subunit(RPC14)
		AT4G38440	RPAP1-like, carboxy-terminal protein(IYO)
		AT1G48410	Stabilizer of iron transporter SufD / Polynucleotidyl transferase(AGO1)
		AT1G04950	TATA BOX ASSOCIATED FACTOR II 59(TAFII59)
		AT1G54140	TATA binding protein associated factor 21kDa subunit(TAFII21)
		AT1G50300	TBP-associated factor 15(TAF15)
		AT5G58470	TBP-associated factor 15B(TAF15b)
		AT5G43130	TBP-associated factor 4(TAF4)
		AT5G25150	TBP-associated factor 5(TAF5)
		AT3G25940	TFIIB zinc-binding protein(AT3G25940)
		AT3G12250	TGACG motif-binding factor 6(TGA6)
		AT5G09850	Transcription elongation factor (TFIIS) family protein(AT5G09850)
		AT1G32130	Transcription elongation factor (TFIIS) family protein(IWS1)
		AT1G63210	Transcription elongation factor Spt6(AT1G63210)
		AT4G12610	transcription initiation factor IIF subunit alpha RAP74(RAP74)
		AT5G08565	Transcription initiation Spt4-like protein(AT5G08565)
		AT4G20330	Transcription initiation factor TFIIE, beta subunit(AT4G20330)
		AT4G35050	Transducin family protein / WD-40 repeat family protein(MSI3)
		AT5G58230	Transducin/WD40 repeat-like superfamily protein(MSI1)
		AT5G59710	VIRE2 interacting protein 2(VIP2)
		AT1G43700	VIRE2-interacting protein 1(VIP1)
		AT1G17880	basic transcription factor 3(BTF3)
		AT2G06210	binding protein(ELF8)
		AT5G14270	bromodomain and extraterminal domain protein 9(BET9)
		AT1G77180	chromatin protein family(SKIP)
		AT3G10010	demeter-like 2(DML2)
		AT5G52470	fibrillarin 1(FIB1)
		AT4G25630	fibrillarin 2(FIB2)
		AT1G05055	general transcription factor II H2(GTF2H2)
		AT1G10270	glutamine-rich protein 23(GRP23)
		AT5G02500	heat shock cognate protein 70-1(HSC70-1)

		AT3G12580	heat shock protein 70(HSP70)
		AT3G44530	histone chaperone HIRA-like protein(HIRA)
		AT4G38130	histone deacetylase 1(HD1)
		AT3G18520	histone deacetylase 15(HDA15)
		AT5G22650	histone deacetylase 2B(HD2B)
		AT5G03740	histone deacetylase 2C(HD2C)
		AT5G26040	histone deacetylase 2(HDA2)
		AT3G44750	histone deacetylase 3(HDA3)
		AT5G63110	histone deacetylase 6(HDA6)
		AT2G27840	histone deacetylase-related / HD-like protein(HDT4)
		AT5G02850	hydroxyproline-rich glycoprotein family protein(AT5G02850)
		AT1G79730	hydroxyproline-rich glycoprotein family protein(ELF7)
		AT5G61150	leo1-like family protein(VIP4)
		AT4G04780	mediator 21(MED21)
		AT1G15780	mediator of RNA polymerase II transcription subunit 15a-like protein(NRB4)
		AT2G03070	mediator subunit 8(MED8)
		AT3G63030	methyl-CPG-binding domain 4(MBD4)
		AT3G46580	methyl-CPG-binding domain protein 5(MBD5)
		AT4G29730	nucleosome/chromatin assembly factor group C5(NFC5)
		AT3G14890	phosphoesterase(AT3G14890)
		AT1G25540	phytochrome and flowering time regulatory protein (PFT1)(PFT1)
		AT1G61040	plus-3 domain-containing protein(VIP5)
		AT1G44910	pre-mRNA-processing protein 40A(PRP40A)
		AT3G19670	pre-mRNA-processing protein 40B(PRP40B)
		AT3G19840	pre-mRNA-processing protein 40C(PRP40C)
		AT2G30120	protein FLC EXPRESSOR(AT2G30120)
		AT2G32080	purin-rich alpha 1(PUR ALPHA-1)
		AT5G67240	small RNA degrading nuclease 3(SDN3)
		AT2G27040	Argonaute family protein(AGO4)
		AT4G12050	Putative AT-hook DNA-binding family protein(AT4G12050)
		AT1G03770	RING 1B(RING1B)
		AT3G10070	TBP-associated factor 12(TAF12)
		AT5G23150	Tudor/PWWP/MBT domain-containing protein(HUA2)
		AT2G35670	VEFS-Box of polycomb protein(FIS2)
		AT5G17690	like heterochromatin protein (LHP1)(TFL2)
		AT1G55080	mediator of RNA polymerase II transcription subunit-like protein(MED9)
		AT3G04510	LIGHT-DEPENDENT SHORT HYPOCOTYLS-like protein (DUF640)(LSH2)
		AT1G26665	Mediator complex, subunit Med10(AT1G26665)
		AT4G31720	TBP-associated factor II 15(TAFII15)
		AT2G36490	demeter-like 1(DML1)
		AT2G34640	plastid transcriptionally active 12(PTAC12)

Supplementary table 2. List of candidate proteins for nuclear import under effect of flg22 and nlp20.

Flg22	
Accession	Description
AT3G55100	ABC-2 type transporter family protein
AT4G33950	Protein kinase superfamily protein
AT3G52050	5'-3' exonuclease family protein
AT5G53340	Galactosyltransferase family protein
AT5G22630	arogenate dehydratase 5
AT4G21160	Calcium-dependent ARF-type GTPase activating protein family
AT1G06510	unknown protein
AT3G26080	plastid-lipid associated protein PAP / fibrillin family protein
AT4G27010	EMB2788, EMBRYO DEFECTIVE 2788, Ribosome 60S biogenesis N-terminal
AT3G45180	Ubiquitin-like superfamily protein
AT1G66460	Protein kinase superfamily protein
AT1G21170	Exocyst complex component SEC5
AT5G44110	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G26890	DNAJ heat shock N-terminal domain-containing protein
AT5G48940	Leucine-rich repeat transmembrane protein kinase family protein
AT5G03100	F-box/RNI-like superfamily protein
AT5G20040	isopentenyltransferase 9
AT5G01350	unknown protein
AT5G06930	nucleolar-like protein;(source:Araport11)
AT1G10940	Protein kinase superfamily protein
AT2G32900	centromere/kinetochore protein, putative (ZW10)
AT5G63880	SNF7 family protein
AT5G57320	villin, putative
AT2G21510	DNAJ heat shock N-terminal domain-containing protein
AT1G54320	LEM3 (ligand-effect modulator 3) family protein / CDC50 family protein
AT2G32120	heat-shock protein 70T-2 ; heat-shock protein 70T-2
AT1G02520	P-glycoprotein 11
AT3G05040	ARM repeat superfamily protein
AT5G13160	Protein kinase superfamily protein
AT2G34560	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G01270	quiescin-sulfhydryl oxidase 2
AT2G19470	casein kinase I-like 5
AT5G30490	craniofacial development-like protein;(source:Araport11)
AT3G27080	translocase of outer membrane 20 kDa subunit 3
AT2G17760	Eukaryotic aspartyl protease family protein

AT2G38310	PYR1-like 4	
AT4G21900	proteinaceous RNase P 3	
AT5G09270	unknown protein, transmembrane protein;(source:Araport11)	
AT5G19620	outer envelope protein of 80 kDa	
AT4G29060	elongation factor Ts family protein	
AT5G13780	Acyl-CoA N-acyltransferases (NAT) superfamily protein	
AT1G13980	sec7 domain-containing protein	
AT2G24290	Protein of unknown function (DUF1068)	
AT1G55080	MED9	
AT3G09560	Lipin family protein ; Lipin family protein ; Lipin family protein	
AT5G42470	BRCA1-A complex subunit BRE-like protein;(source:Araport11)	
AT3G54050	high cyclic electron flow 1	
AT5G47020	MraZ;(source:Araport11)	
AT5G23150	Tudor/PWWP/MBT domain-containing protein	
AT5G37370	PRP38 family protein	
AT4G11860	Protein of unknown function (DUF544)	
AT5G40930	translocase of outer membrane 20-4	
AT5G47650	nudix hydrolase homolog 2	
AT1G71400	receptor like protein 12	
AT5G39500	GNOM-like 1	
AT4G32680	G1IP, GET1-INTERACTING PROTEIN, GET2	
AT5G24840	tRNA (guanine-N-7) methyltransferase	
AT1G68300	Adenine nucleotide alpha hydrolases-like superfamily protein	
AT3G05100	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	
AT5G23630	phosphate deficiency response 2	
AT1G50460	hexokinase-like 1	
AT2G30740	Protein kinase superfamily protein	
AT3G11070	Outer membrane OMP85 family protein	
AT3G17910	Surfeit locus 1 cytochrome c oxidase biogenesis protein	
AT1G77140	vacuolar protein sorting 45	
AT1G30480	D111/G-patch domain-containing protein	
AT2G44525	Protein of unknown function (DUF498/DUF598)	
AT4G26430	COP9 signalosome subunit 6B	
AT1G65540	LETM1-like protein	
AT1G12230	Aldolase superfamily protein	
AT4G16450	unknown protein; NADH-ubiquinone oxidoreductase;(source:Araport11)	
AT2G22475	GRAM domain family protein	
AT5G20350	Ankyrin repeat family protein with DHHC zinc finger domain	
AT3G02540	Rad23 UV excision repair protein family	
AT1G79650	Rad23 UV excision repair protein family	
AT4G30600	signal recognition particle receptor alpha subunit family protein	

AT2G38020	vacuoleless1 (VCL1)
AT5G14680	Adenine nucleotide alpha hydrolases-like superfamily protein
AT4G39980	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1
AT1G06840	Leucine-rich repeat protein kinase family protein
AT5G15910	NAD(P)-binding Rossmann-fold superfamily protein
AT1G21370	unknown protein; transmembrane protein;(source:Araport11)
AT1G66680	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT1G27980	dihydrosphingosine phosphate lyase
AT1G52530	Hus1-like protein;(source:Araport11)
AT1G54570	Esterase/lipase/thioesterase family protein
AT5G59520	ZRT/IRT-like protein 2
AT4G34290	SWIB/MDM2 domain superfamily protein
AT5G05380	prenylated RAB acceptor 1.B3
AT5G08440	Unknown protein, transmembrane protein;(source:Araport11)
AT2G28390	SAND family protein
AT5G04990	SAD1/UNC-84 domain protein 1
AT1G18270	ketose-bisphosphate aldolase class-II family protein
AT2G27040	Argonaute family protein
AT5G23880	cleavage and polyadenylation specificity factor 100
AT2G44680	casein kinase II beta subunit 4
AT4G24290	MAC/Perforin domain-containing protein
AT3G01520	Adenine nucleotide alpha hydrolases-like superfamily protein
AT4G14070	acyl-activating enzyme 15
AT3G44340	clone eighty-four
AT1G02130	RAS 5
AT2G44650	chloroplast chaperonin 10
AT5G46210	cullin4
AT2G37970	SOUL heme-binding family protein
AT5G06140	sorting nexin 1
AT1G55810	uridine kinase-like 3 ; uridine kinase-like 3
AT2G38550	Transmembrane proteins 14C
AT1G22200	Endoplasmic reticulum vesicle transporter protein
AT1G28340	receptor like protein 4
AT5G15400	U-box domain-containing protein
AT3G59020	ARM repeat superfamily protein
AT2G34590	Transketolase family protein
AT3G13870	Root hair defective 3 GTP-binding protein (RHD3)
AT2G22480	phosphofructokinase 5
AT1G36390	Co-chaperone GrpE family protein
AT5G05520	Outer membrane OMP85 family protein
AT1G05350	NAD(P)-binding Rossmann-fold superfamily protein

AT1G60660	cytochrome B5-like protein
AT4G17530	RAB GTPase homolog 1C
AT5G59150	RAB GTPase homolog A2D
AT3G43520	Transmembrane proteins 14C
AT5G10450	G-box regulating factor 6
AT2G21060	glycine-rich protein 2B
AT3G58140	phenylalanyl-tRNA synthetase class IIc family protein
AT5G11490	adaptin family protein
AT5G27600	long-chain acyl-CoA synthetase 7
AT5G62700	tubulin beta chain 3 ; tubulin beta chain 2
AT1G20200	PAM domain (PCI/PINT associated module) protein
AT2G29550	tubulin beta-7 chain
AT1G30690	Sec14p-like phosphatidylinositol transfer family protein
AT3G20630	ubiquitin-specific protease 14
AT4G17870	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT4G00570	NAD-dependent malic enzyme 2
AT1G11660	heat shock protein 70 (Hsp 70) family protein
AT5G03520	RAB GTPase homolog 8C
AT2G45290	Transketolase
AT3G44300	nitrilase 2
AT4G30010	ATP-dependent RNA helicase;(source:Araport11)
AT4G20400	JUMONJI 14
AT4G22240	Plastid-lipid associated protein PAP / fibrillin family protein
AT5G16390	chloroplastic acetylcoenzyme A carboxylase 1
AT5G14105	unknown protein
AT5G58700	phosphatidylinositol-speciwc phospholipase C4
AT1G29790	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT3G59500	Integral membrane HRF1 family protein
AT5G04260	WCRKC thioredoxin 2
AT4G16660	heat shock protein 70 (Hsp 70) family protein
AT4G31985	Ribosomal protein L39 family protein
AT4G20360	RAB GTPase homolog E1B
AT3G43300	HOPM interactor 7
AT3G02760	Class II aaRS and biotin synthetases superfamily protein
AT2G47970	Nuclear pore localisation protein NPL4
AT5G53480	ARM repeat superfamily protein
AT5G25450	Cytochrome bd ubiquinol oxidase, 14kDa subunit
AT2G17265	homoserine kinase
AT5G08530	51 kDa subunit of complex I
AT3G20000	translocase of the outer mitochondrial membrane 40

Nlp20	
Accession	Description
AT2G39390	Ribosomal L29 family protein
AT3G43980	Ribosomal protein S14p/S29e family protein
AT4G14360	S-adenosyl-L-methionine-dependent methyltransferases superfamily
AT3G55750	Ribosomal protein L35Ae family protein
AT4G38780	Pre-mRNA-processing-splicing factor
AT2G32670	vesicle-associated membrane protein 725
AT3G60810	DUF1499 family protein;(source:Araport11)
AT1G53070	Legume lectin family protein
AT1G64650	Major facilitator superfamily protein
AT5G49830	exocyst complex component 84B
AT5G37020	auxin response factor 8
AT2G38440	SCAR homolog 2
AT5G58510	Rab3 GTPase-activating protein catalytic protein;(source:Araport11)
AT4G38040	Exostosin family protein
AT1G52300	Zinc-binding ribosomal protein family protein
AT5G42100	beta-1,3-glucanase_putative
AT5G64410	oligopeptide transporter 4
AT1G33940	Serine/Threonine-kinase ULK4-like protein;(source:Araport11)
AT3G53020	Ribosomal protein L24e family protein
AT4G14455	Target SNARE coiled-coil domain protein
AT3G16860	COBRA-like protein 8 precursor
AT5G06410	DNAJ heat shock N-terminal domain-containing protein
AT1G20260	ATPase, V1 complex, subunit B protein
AT3G18410	Complex I subunit NDUFS6 ; Complex I subunit NDUFS6
AT2G19830	SNF7 family protein
AT3G18420	Protein prenyltransferase superfamily protein
AT4G29660	embryo defective 2752
AT1G31760	SWIB/MDM2 domain superfamily protein
AT1G29260	peroxin 7
AT4G03550	glucan synthase-like 5
AT3G24420	alpha/beta-Hydrolases superfamily protein
AT3G56210	ARM repeat superfamily protein
AT4G36945	PLC-like phosphodiesterases superfamily protein
AT3G25290	Auxin-responsive family protein
AT4G00026	SD3, SEGREGATION DISTORTION 3
AT1G71840	transducin family protein / WD-40 repeat family protein
AT2G02500	Nucleotide-diphospho-sugar transferases superfamily protein
AT3G04820	Pseudouridine synthase family protein

AT3G49120	peroxidase CB	
AT1G09630	RAB GTPase 11C	
AT5G22120	coiled-coil protein;(source:Araport11)	
AT2G20900	diacylglycerol kinase 5 ; diacylglycerol kinase 5	
AT5G54190	protochlorophyllide oxidoreductase A	
AT1G16810	7-dehydrocholesterol reductase-like protein;(source:Araport11)	
AT1G05720	selenoprotein family protein	
AT3G04950	SEC-C motif protein;(source:Araport11)	
AT1G60740	Thioredoxin superfamily protein	
AT4G02450	HSP20-like chaperones superfamily protein	
AT5G15120	Protein of unknown function (DUF1637)	
AT1G49880	Erv1/Alr family protein	
AT1G79010	Alpha-helical ferredoxin	
AT5G60790	ABC transporter family protein	
AT4G02520	glutathione S-transferase PHI 2	
AT2G28000	chaperonin-60alpha	
AT5G38830	Cysteinyl-tRNA synthetase, class Ia family protein	
AT3G44190	FAD/NAD(P)-binding oxidoreductase family protein	
AT4G31180	Class II aminoacyl-tRNA and biotin synthetases superfamily protein	
AT5G59320	lipid transfer protein 3	
AT2G44310	Calcium-binding EF-hand family protein	
AT5G64290	dicarboxylate transport 2.1	
AT3G18430	Calcium-binding EF-hand family protein	
AT3G08510	phospholipase C 2 ; phospholipase C 2	
AT4G31500	cytochrome P450, family 83, subfamily B, polypeptide 1	
AT1G13560	aminoalcoholphosphotransferase 1	
AT4G32520	serine hydroxymethyltransferase 3	
AT1G56050	GTP-binding protein-related	
AT1G15270	Translation machinery associated TMA7	
AT1G35220	FAM91 carboxy-terminus protein;(source:Araport11)	
AT5G17990	tryptophan biosynthesis 1	
AT5G14150	Protein of unknown function, DUF642	
AT1G74310	heat shock protein 101	
AT2G25070	Protein phosphatase 2C family protein	
AT2G16600	rotamase CYP 3	

Supplementary table 3. List of proteins with significant change in abundance between the conditions.

Multiple test (ANOVA), FDR: 0.05				F: flg22, N: Nlp20, C: control	Significant pairs, X_Y	X= F or N or C
					X= high, Y= Low	Y= F or N or C
	ANOVA Significant	-Log ANOVA p value	ANOVA q-value		Significant pairs	
AT3G20670	+	6.57657	0		F_C;F_N	
AT1G26910	+	6.58291	0		F_C;F_N	
AT5G20000	+	9.3903	0		F_C;F_N	
AT5G13780	+	3.27737	0.010643		F_C;F_N	
AT2G44525	+	4.14164	0.002171		F_C;F_N	
AT1G65540	+	2.67344	0.0327		F_C;F_N	
AT5G14680	+	3.31127	0.009964		F_C;F_N	
AT4G27270	+	3.32253	0.009926		F_C;F_N	
AT2G34590	+	3.71327	0.005067		C_N;F_N	
AT1G20200	+	2.51896	0.042212		F_C;F_N	
AT5G03520	+	7.98074	0		F_C;F_N	
AT3G02760	+	3.83124	0.004		F_C;F_N	
AT5G64650	+	5.9258	0		N_C;N_F	
AT4G18430	+	8.55696	0		N_C;N_F	
AT2G39390	+	6.90961	0		N_C;N_F	
AT4G38780	+	8.38945	0		N_C;N_F	
AT1G17730	+	4.32893	0.001647		N_C;N_F	
AT4G27680	+	7.31322	0		N_C;N_F	
AT1G04480	+	6.10488	0		N_C;N_F	
AT4G13170	+	6.857	0		N_C;N_F	
AT3G53020	+	7.28204	0		N_C;N_F	
AT1G20260	+	12.8549	0		N_C;N_F	
AT1G09630	+	11.3538	0		N_C;N_F	
AT4G02520	+	4.42684	0.0015		N_C;N_F	
AT5G09500	+	5.49738	0.000174		C_F;C_N	
AT5G59240	+	2.9377	0.020182		C_F;C_N	
AT1G20010	+	3.69783	0.00513		C_N;F_N	
AT2G40010	+	7.41618	0		C_F;C_N	
AT3G22310	+	2.57895	0.038169		C_F;C_N	
AT3G28900	+	3.1681	0.013533		C_F;C_N	
AT2G46610	+	2.90374	0.021851		C_F;C_N	
AT2G40590	+	2.76839	0.028487		C_F;C_N	
AT1G09100	+	7.81788	0		C_F;C_N	
AT2G40510	+	2.7945	0.027111		C_F;C_N	
AT5G15520	+	4.49126	0.001333		C_F;C_N	

AT4G11840	+	4.45022	0.001571		C_F;C_N		
AT2G36410	+	3.32896	0.009962		C_F;C_N		
AT2G29200	+	2.76972	0.028767		C_F;C_N		
AT1G07930	+	7.32754	0		C_F;C_N		
AT5G58420	+	4.01716	0.002564		C_F;C_N		
AT5G43070	+	4.44705	0.001467		C_F;C_N		
AT1G54270	+	2.44833	0.045807		N_F;C_F		
AT3G11500	+	4.09626	0.002667		C_F;C_N		
AT5G18380	+	4.64999	0.00112		C_F;C_N		
AT1G50200	+	2.64527	0.033902		F_N		
AT2G28760	+	8.49665	0		C_F;C_N		
AT2G20530	+	7.03008	0		N_C;F_C		
AT5G42540	+	2.49115	0.043727		F_C;F_N		
AT2G23930	+	3.7326	0.005302		N_C;F_C		
AT4G30800	+	3.04566	0.017419		N_C;F_C		
AT2G21410	+	4.4062	0.001455		N_C;F_C		
AT3G48820	+	3.79575	0.004286		N_C;F_C;F_N		
AT5G07090	+	4.53538	0.001231		N_C;F_C		
AT3G22230	+	3.27233	0.010807		N_C;F_C		
AT2G09990	+	2.83077	0.026087		N_C;F_C		
AT5G23740	+	2.57096	0.038381		F_C		
AT1G16030	+	5.99422	0		N_C;F_C		
AT4G08520	+	3.15713	0.013836		F_C;F_N		
AT3G06860	+	3.4647	0.007373		N_C;F_C;F_N		
AT4G16120	+	3.55248	0.006449		N_C;N_F		
AT5G12290	+	3.69213	0.005021		N_C;F_C		
AT3G25800	+	2.95485	0.020444		F_C		
AT2G31410	+	4.4434	0.001419		C_F;C_N		
AT5G54600	+	2.70153	0.031392		C_F		
AT5G25150	+	2.49026	0.043326		C_F;N_F		
AT3G60180	+	2.45615	0.045522		C_F;N_F		
AT1G78050	+	2.71856	0.031474		C_F;N_F		
AT2G20630	+	4.44718	0.001517		C_F;N_F;N_C		
AT1G14320	+	2.49296	0.044092		C_F;N_F		
AT1G29970	+	2.81146	0.026971		C_F;N_F		
AT2G36620	+	8.19067	0		F_N;C_N		
AT3G02320	+	4.0773	0.002703		F_N;F_C		
AT5G56000	+	3.21903	0.012		C_N;F_N		
AT3G04400	+	3.39667	0.008692		C_N;F_N		
AT3G27230	+	4.01012	0.0026		F_N;F_C		
AT3G48670	+	2.71747	0.031221		F_N;F_C		

AT2G47650	+	4.07152	0.002632		C_N;F_N		
AT1G21190	+	2.95365	0.020125		C_N;F_N		
AT3G53650	+	2.79982	0.027268		N_C;N_F		
AT3G06610	+	2.87308	0.023882		C_N;C_F		
AT5G21160	+	2.65363	0.033926		C_N		
AT3G03960	+	2.46475	0.045099		F_C		
AT4G28510	+	4.85309	0.000667		N_C;F_C		
AT3G02090	+	3.59001	0.006333		N_C;F_C		
AT5G48370	+	3.52844	0.00664		F_N;F_C		
AT1G03860	+	6.1608	0		N_C;F_C		
AT5G15090	+	2.47006	0.045067		F_C;N_C		
AT3G15730	+	2.51445	0.042093		N_C;F_C		
AT3G27240	+	2.71342	0.031077		F_C;F_N		
AT3G27280	+	3.25833	0.011035		F_C;N_C		
AT5G40810	+	2.73219	0.03072		F_C;F_N		
AT4G39280	+	2.94074	0.020369		F_C;F_N		
AT5G40770	+	3.7149	0.005182		F_C;N_C		

Supplementary table 4. Proteins annotated to the transcription process were classified into families. UP_KEYWORDS category was used in the David bioinformatics resources annotation

Transcription Factors families	Number of proteins		
AP2	1	AT2G41710	Integrase-type DNA-binding superfamily protein(AT2G41710)
ARF	2	AT5G20730	Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-like protein(NPH4)
		AT5G62000	auxin response factor 2(ARF2)
ARR-B	2	AT3G16857	response regulator 1(RR1)
		AT4G16110	response regulator 2(RR2)
B3	2	AT1G49480	related to vernalization1 1(RTV1)
		AT3G18990	AP2/B3-like transcriptional factor family protein(VRN1)
BBR-BPC	4	AT1G14685	basic pentacysteine 2(BPC2)
		AT2G01930	basic pentacysteine1(BPC1)
		AT2G21240	basic pentacysteine 4(BPC4)
		AT5G42520	basic pentacysteine 6(BPC6)
BES1	1	AT1G75080	Brassinosteroid signaling positive regulator (BZR1) family protein(BZR1)
bHLH	19	AT1G01260	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G01260)
		AT1G03040	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G03040)
		AT1G05710	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G05710)
		AT1G05805	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G05805)
		AT1G27660	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G27660)

		AT1G35460	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH1)
		AT2G31730	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT2G31730)
		AT2G42280	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH4)
		AT2G43140	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT2G43140)
		AT2G46510	ABA-inducible BHLH-type transcription factor(AIB)
		AT3G19860	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(bHLH121)
		AT3G20640	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT3G20640)
		AT3G59060	phytochrome interacting factor 3-like 6(PIL6)
		AT4G02590	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(UNE12)
		AT4G09180	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(FBH2)
		AT4G16430	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT4G16430)
		AT5G08130	basic helix-loop-helix (bHLH) DNA-binding superfamily protein(BIM1)
		AT5G38860	BES1-interacting Myc-like protein 3(BIM3)
		AT5G46760	Basic helix-loop-helix (bHLH) DNA-binding family protein(MYC3)
bZIP	16	AT1G03970	G-box binding factor 4(GBF4)
		AT1G22070	transcription factor TGA3(TGA3)
		AT1G32150	basic region/leucine zipper transcription factor 68(bZIP68)
		AT1G43700	VIRE2-interacting protein 1(VIP1)
		AT1G45249	abscisic acid responsive elements-binding factor 2(ABF2)
		AT1G77920	bZIP transcription factor family protein(TGA7)
		AT2G21230	Basic-leucine zipper (bZIP) transcription factor family protein(AT2G21230)
		AT2G31370	Basic-leucine zipper (bZIP) transcription factor family protein(AT2G31370)
		AT2G35530	basic region/leucine zipper transcription factor 16(bZIP16)
		AT3G12250	TGACG motif-binding factor 6(TGA6)
		AT3G19290	ABRE binding factor 4(ABF4)
		AT3G54620	basic leucine zipper 25(BZIP25)
		AT3G56850	ABA-responsive element binding protein 3(AREB3)
		AT4G35040	Basic-leucine zipper (bZIP) transcription factor family protein(bZIP19)
		AT4G36730	G-box binding factor 1(GBF1)
		AT4G38900	Basic-leucine zipper (bZIP) transcription factor family protein(AT4G38900)
C2H2	10	AT1G14580	C2H2-like zinc finger protein(AT1G14580)
		AT2G02070	indeterminate(ID)-domain 5(IDD5)
		AT2G02080	indeterminate(ID)-domain 4(IDD4)
		AT2G41940	zinc finger protein 8(ZFP8)
		AT3G13810	indeterminate(ID)-domain 11(IDD11)
		AT3G19580	zinc-finger protein 2(ZF2)
		AT3G45260	C2H2-like zinc finger protein(AT3G45260)
		AT3G50700	indeterminate(ID)-domain 2(IDD2)
		AT4G06634	zinc finger (C2H2 type) family protein(YY1)
		AT5G66730	C2H2-like zinc finger protein(IDD1)
C3H	1	AT2G16485	GW repeat- and PHD finger-containing protein NERD(NERD)

CAMTA	2	AT2G22300	signal responsive 1(SR1)
		AT5G64220	Calmodulin-binding transcription activator protein with CG-1 and Ankyrin domain(AT5G64220)
Dof	7	AT1G64620	Dof-type zinc finger DNA-binding family protein(AT1G64620)
		AT2G46590	Dof-type zinc finger DNA-binding family protein(DAG2)
		AT3G21270	DOF zinc finger protein 2(DOF2)
		AT5G02460	Dof-type zinc finger DNA-binding family protein(AT5G02460)
		AT5G39660	cycling DOF factor 2(CDF2)
		AT5G60850	OBF binding protein 4(OBP4)
		AT5G65590	Dof-type zinc finger DNA-binding family protein(AT5G65590)
E2F/DP	2	AT5G03415	Transcription factor DP(DPB)
		AT5G22220	E2F transcription factor 1(E2F1)
ERF	2	AT2G44840	ethylene-responsive element binding factor 13(ERF13)
		AT5G13330	related to AP2 6l(Rap2.6L)
G2-like	10	AT1G25550	myb-like transcription factor family protein(AT1G25550)
		AT1G79430	Homeodomain-like superfamily protein(APL)
		AT2G20400	myb-like HTH transcriptional regulator family protein(AT2G20400)
		AT3G04030	Homeodomain-like superfamily protein(MYR2)
		AT3G12730	Homeodomain-like superfamily protein(AT3G12730)
		AT3G13040	myb-like HTH transcriptional regulator family protein(AT3G13040)
		AT4G28610	phosphate starvation response 1(PHR1)
		AT5G05090	Homeodomain-like superfamily protein(AT5G05090)
		AT5G18240	myb-related protein 1(MYR1)
		AT5G29000	Homeodomain-like superfamily protein(PHL1)
GATA	2	AT1G51600	ZIM-LIKE 2(ZML2)
		AT5G47140	GATA transcription factor 27(GATA27)
GeBP	4	AT1G61730	DNA-binding storekeeper protein-related transcriptional regulator(AT1G61730)
		AT3G04930	DNA-binding storekeeper protein-related transcriptional regulator(AT3G04930)
		AT4G00270	DNA-binding storekeeper protein-related transcriptional regulator(AT4G00270)
		AT5G28040	DNA-binding storekeeper protein-related transcriptional regulator(AT5G28040)
GRAS	4	AT1G07530	SCARECROW-like 14(SCL14)
		AT1G21450	SCARECROW-like 1(SCL1)
		AT1G50600	scarecrow-like 5(SCL5)
		AT4G17230	SCARECROW-like 13(SCL13)
HB-other	2	AT1G28420	homeobox-1(HB-1)
		AT5G44180	Homeodomain-like transcriptional regulator(RLT2)
HB-PHD	1	AT3G19510	Homeodomain-like protein with RING/FYVE/PHD-type zinc finger domain-containing protein(HAT3.1)
HD-ZIP	3	AT4G00730	Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein(ANL2)
		AT4G04890	protodermal factor 2(PDF2)
		AT5G65310	homeobox protein 5(HB5)
HSF	8	AT1G32330	heat shock transcription factor A1D(HSFA1D)

		AT1G67970	heat shock transcription factor A8(HSFA8)
		AT3G24520	heat shock transcription factor C1(HSFC1)
		AT4G11660	winged-helix DNA-binding transcription factor family protein(AT-HSFB2B)
		AT4G13980	winged-helix DNA-binding transcription factor family protein(AT-HSFA5)
		AT4G17750	heat shock factor 1(HSF1)
		AT4G18880	heat shock transcription factor A4A(HSF A4A)
		AT5G16820	heat shock factor 3(HSF3)
MYB	5	AT1G09770	cell division cycle 5(CDC5)
		AT3G09370	myb domain protein 3r-3(MYB3R-3)
		AT3G52250	Duplicated homeodomain-like superfamily protein(AT3G52250)
		AT4G32730	Homeodomain-like protein(PC-MYB1)
		AT5G08520	Duplicated homeodomain-like superfamily protein(AT5G08520)
MYB_related	7	AT1G49950	telomere repeat binding factor 1(TRB1)
		AT1G72740	Homeodomain-like/winged-helix DNA-binding family protein(AT1G72740)
		AT2G47210	myb-like transcription factor family protein(AT2G47210)
		AT3G09600	Homeodomain-like superfamily protein(RVE8)
		AT4G01280	Homeodomain-like superfamily protein(AT4G01280)
		AT5G47390	myb-like transcription factor family protein(AT5G47390)
		AT5G67580	Homeodomain-like/winged-helix DNA-binding family protein(TRB2)
NAC	7	AT1G25580	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein(SOG1)
		AT1G33060	NAC 014(NAC014)
		AT1G52890	NAC domain containing protein 19(NAC019)
		AT3G10480	NAC domain containing protein 50(NAC050)
		AT3G10490	NAC domain containing protein 52(NAC052)
		AT3G15500	NAC domain containing protein 3(NAC3)
		AT4G35580	NAC transcription factor-like 9(NTL9)
NF-YC	2	AT1G56170	nuclear factor Y, subunit C2(NF-YC2)
		AT3G48590	nuclear factor Y, subunit C1(NF-YC1)
Nin-like	1	AT4G35270	Plant regulator RWP-RK family protein(AT4G35270)
SBP	2	AT3G15270	squamosa promoter binding protein-like 5(SPL5)
		AT5G50670	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein(SPL13B)
TALE	5	AT1G19700	BEL1-like homeodomain 10(BEL10)
		AT1G75410	BEL1-like homeodomain 3(BLH3)
		AT2G23760	BEL1-like homeodomain 4(BLH4)
		AT2G35940	BEL1-like homeodomain 1(BLH1)
		AT4G36870	BEL1-like homeodomain 2(BLH2)
TCP	8	AT1G35560	TCP family transcription factor(AT1G35560)
		AT1G53230	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 3(TCP3)
		AT1G58100	TCP family transcription factor(TCP8)
		AT1G72010	TCP family transcription factor(AT1G72010)
		AT2G31070	TCP domain protein 10(TCP10)

		AT2G45680	TCP family transcription factor(TCP9)
		AT3G02150	plastid transcription factor 1(PTF1)
		AT3G47620	TEOSINTE BRANCHED, cycloidea and PCF (TCP) 14(TCP14)
Trihelix	11	AT1G13450	Homeodomain-like superfamily protein(GT-1)
		AT1G33240	GT-2-like 1(GTL1)
		AT1G54060	6B-interacting protein 1-like 1(ASIL1)
		AT1G76880	Duplicated homeodomain-like superfamily protein(AT1G76880)
		AT1G76890	Duplicated homeodomain-like superfamily protein(GT2)
		AT2G38250	Homeodomain-like superfamily protein(AT2G38250)
		AT3G11100	sequence-specific DNA binding transcription factor(AT3G11100)
		AT3G14180	sequence-specific DNA binding transcription factor(ASIL2)
		AT3G25990	Homeodomain-like superfamily protein(AT3G25990)
		AT5G05550	sequence-specific DNA binding transcription factor(AT5G05550)
		AT5G28300	Duplicated homeodomain-like superfamily protein(AT5G28300)
VOZ	1	AT2G42400	vascular plant one zinc finger protein 2(VOZ2)
WOX	1	AT4G35550	WUSCHEL related homeobox 13(WOX13)
WRKY	19	AT1G13960	WRKY DNA-binding protein 4(WRKY4)
		AT1G18860	WRKY DNA-binding protein 61(WRKY61)
		AT1G62300	WRKY family transcription factor(WRKY6)
		AT1G69310	WRKY DNA-binding protein 57(WRKY57)
		AT1G80840	WRKY DNA-binding protein 40(WRKY40)
		AT2G03340	WRKY DNA-binding protein 3(WRKY3)
		AT2G04880	zinc-dependent activator protein-1(ZAP1)
		AT2G23320	WRKY DNA-binding protein 15(WRKY15)
		AT2G30590	WRKY DNA-binding protein 21(WRKY21)
		AT2G38470	WRKY DNA-binding protein 33(WRKY33)
		AT2G40740	WRKY DNA-binding protein 55(WRKY55)
		AT3G01970	WRKY DNA-binding protein 45(WRKY45)
		AT4G01720	WRKY family transcription factor(WRKY47)
		AT4G26640	WRKY family transcription factor family protein(WRKY20)
		AT4G31550	WRKY DNA-binding protein 11(WRKY11)
		AT4G31800	WRKY DNA-binding protein 18(WRKY18)
		AT5G07100	WRKY DNA-binding protein 26(WRKY26)
		AT5G24110	WRKY DNA-binding protein 30(WRKY30)
		AT5G56270	WRKY DNA-binding protein 2(WRKY2)
ZF-HD	3	AT3G28920	homeobox protein 34(HB34)
		AT5G15210	homeobox protein 30(HB30)
		AT5G65410	homeobox protein 25(HB25)
Others	Number of proteins		
	145	AT4G16420	ADA2 2B(ADA2B)

		AT1G20910	ARID/BRIGHT DNA-binding domain-containing protein(AT1G20910)
		AT1G76510	ARID/BRIGHT DNA-binding domain-containing protein(AT1G76510)
		AT2G17410	ARID/BRIGHT DNA-binding domain-containing protein(AT2G17410)
		AT2G33620	AT hook motif DNA-binding family protein(AT2G33620)
		AT3G04590	AT hook motif DNA-binding family protein(AT3G04590)
		AT4G25320	AT hook motif DNA-binding family protein(AT4G25320)
		AT4G12080	AT-hook motif nuclear-localized protein 1(AHL1)
		AT1G08060	ATP-dependent helicase family protein(MOM)
		AT1G17450	B-block binding subunit of TFIIIC(AT1G17450)
		AT3G42170	BED zinc finger and hAT dimerization domain-containing protein DAYSLEEPER(DAYSLEEPER)
		AT1G18560	BED zinc finger and hAT dimerization domain-containing protein(AT1G18560)
		AT1G55750	BSD domain (BTF2-like transcription factors, Synapse-associated proteins and DOS2-like proteins)(AT1G55750)
		AT4G21670	C-terminal domain phosphatase-like 1(CPL1)
		AT2G33540	C-terminal domain phosphatase-like 3(CPL3)
		AT4G27430	COP1-interacting protein 7(CIP7)
		AT5G57580	Calmodulin-binding protein(AT5G57580)
		AT3G09360	Cyclin/Brf1-like TBP-binding protein(AT3G09360)
		AT1G18950	DDT domain superfamily(AT1G18950)
		AT3G48710	DEK domain-containing chromatin associated protein(AT3G48710)
		AT4G26630	DEK domain-containing chromatin associated protein(AT4G26630)
		AT5G55660	DEK domain-containing chromatin associated protein(AT5G55660)
		AT5G41880	DNA primase POLA3(POLA3)
		AT4G34430	DNA-binding family protein(CHB3)
		AT4G08540	DNA-directed RNA polymerase II protein(AT4G08540)
		AT5G10110	DNA-directed RNA polymerase subunit beta(AT5G10110)
		AT2G36530	Enolase(LOS2)
		AT5G08550	GC-rich sequence DNA-binding factor-like protein(ILP1)
		AT1G32750	HAC13 protein (HAC13)(HAF01)
		AT5G42020	Heat shock protein 70 (Hsp 70) family protein(BIP2)
		AT5G02490	Heat shock protein 70 (Hsp 70) family protein(Hsp70-2)
		AT4G20400	JUMONJI 14(JMJ14)
		AT2G32700	LEUNIG-like protein(LUH)
		AT4G16310	LSD1-like 3(LDL3)
		AT1G26665	Mediator complex, subunit Med10(AT1G26665)
		AT5G41910	Mediator complex, subunit Med10(MED10A)
		AT1G08600	P-loop containing nucleoside triphosphate hydrolases superfamily protein(ATRX)
		AT5G22330	P-loop containing nucleoside triphosphate hydrolases superfamily protein(RIN1)
		AT2G28290	P-loop containing nucleoside triphosphate hydrolases superfamily protein(SYD)
		AT1G54390	PHD finger protein-like protein(ING2)
		AT3G22590	PLANT HOMOLOGOUS TO PARAFIBROMIN(PHP)

		AT4G28910	Putative interactor of JAZ(NINJA)
		AT5G13010	RNA helicase family protein(EMB3011)
		AT4G35800	RNA polymerase II large subunit(NRPB1)
		AT1G71080	RNA polymerase II transcription elongation factor(AT1G71080)
		AT5G38050	RNA polymerase II transcription elongation factor(AT5G38050)
		AT3G04740	RNA polymerase II transcription mediator(SWP)
		AT4G15950	RNA polymerase II, Rpb4, core protein(NRPD4)
		AT4G25180	RNA polymerase III RPC4(AT4G25180)
		AT4G00830	RNA-binding (RRM/RBD/RNP motifs) family protein(LIF2)
		AT3G04610	RNA-binding KH domain-containing protein(FLK)
		AT4G26000	RNA-binding KH domain-containing protein(PEP)
		AT5G46250	RNA-binding protein(LARP6a)
		AT5G58575	SAGA-associated factor-like protein(AT5G58575)
		AT4G02020	SET domain-containing protein(SWN)
		AT1G43850	SEUSS transcriptional co-regulator(SEU)
		AT5G62090	SEUSS-like 2(SLK2)
		AT5G15020	SIN3-like 2(SNL2)
		AT1G24190	SIN3-like 3(SNL3)
		AT1G70060	SIN3-like 4(SNL4)
		AT1G59890	SIN3-like 5(SNL5)
		AT1G50410	SNF2 domain-containing protein / helicase domain-containing protein / zinc finger protein-like protein(AT1G50410)
		AT5G55100	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein(AT5G55100)
		AT5G14170	SWIB/MDM2 domain superfamily protein(CHC1)
		AT1G16430	Surfeit locus protein 5 subunit 22 of Mediator complex(AT1G16430)
		AT1G07950	Surfeit locus protein 5 subunit 22 of Mediator complex(MED22B)
		AT1G54140	TATA binding protein associated factor 21kDa subunit(TAFII21)
		AT3G10070	TBP-associated factor 12(TAF12)
		AT5G58470	TBP-associated factor 15B(TAF15b)
		AT1G73960	TBP-associated factor 2(TAF2)
		AT5G43130	TBP-associated factor 4(TAF4)
		AT1G27720	TBP-associated factor 4B(TAF4B)
		AT1G55300	TBP-associated factor 7(TAF7)
		AT3G25940	TFIIB zinc-binding protein(AT3G25940)
		AT4G32570	TIFY domain protein 8(TIFY8)
		AT3G16830	TOPLESS-related 2(TPR2)
		AT5G09850	Transcription elongation factor (TFIIS) family protein(AT5G09850)
		AT1G32130	Transcription elongation factor (TFIIS) family protein(IWS1)
		AT2G34900	Transcription factor GTE6(IMB1)
		AT5G24450	Transcription factor IIIC, subunit 5(AT5G24450)
		AT1G03280	Transcription factor TFIIE, alpha subunit(AT1G03280)
		AT4G20340	Transcription factor TFIIE, alpha subunit(AT4G20340)

		AT1G30810	Transcription factor jumonji (jmj) family protein / zinc finger (C5HC2 type) family protein(JMJ18)
		AT1G75510	Transcription initiation factor IIF, beta subunit(AT1G75510)
		AT1G17440	Transcription initiation factor TFIID subunit A(EER4)
		AT2G19520	Transducin family protein / WD-40 repeat family protein(FVE)
		AT2G48160	Tudor/PWWP/MBT domain-containing protein(AT2G48160)
		AT5G23150	Tudor/PWWP/MBT domain-containing protein(HUA2)
		AT3G09670	Tudor/PWWP/MBT superfamily protein(AT3G09670)
		AT5G27650	Tudor/PWWP/MBT superfamily protein(AT5G27650)
		AT5G40340	Tudor/PWWP/MBT superfamily protein(AT5G40340)
		AT3G54500	agglutinin-like protein(AT3G54500)
		AT3G11200	alfin-like 2(AL2)
		AT3G42790	alfin-like 3(AL3)
		AT5G26210	alfin-like 4(AL4)
		AT2G02470	alfin-like 6(AL6)
		AT1G14510	alfin-like 7(AL7)
		AT5G14270	bromodomain and extraterminal domain protein 9(BET9)
		AT5G01270	carboxyl-terminal domain (ctd) phosphatase-like 2(CPL2)
		AT1G77180	chromatin protein family(SKIP)
		AT1G14740	class I heat shock protein, putative (DUF1423)(TTA1)
		AT2G37678	far-red elongated hypocotyl 1(FHY1)
		AT3G63500	fibronectin type III domain protein (DUF1423)(TTA2)
		AT1G73150	global transcription factor group E3(GTE3)
		AT1G06230	global transcription factor group E4(GTE4)
		AT3G27260	global transcription factor group E8(GTE8)
		AT5G02500	heat shock cognate protein 70-1(HSC70-1)
		AT3G28730	high mobility group(HMG)
		AT1G79000	histone acetyltransferase of the CBP family 1(HAC1)
		AT1G16710	histone acetyltransferase of the CBP family 12(HAC12)
		AT4G14385	histone acetyltransferase subunit NuA4-domain protein(AT4G14385)
		AT4G38130	histone deacetylase 1(HD1)
		AT1G79730	hydroxyproline-rich glycoprotein family protein(ELF7)
		AT1G51950	indole-3-acetic acid inducible 18(IAA18)
		AT1G17380	jasmonate-zim-domain protein 5(JAZ5)
		AT5G61150	leo1-like family protein(VIP4)
		AT1G15780	mediator of RNA polymerase II transcription subunit 15a-like protein(NRB4)
		AT1G55080	mediator of RNA polymerase II transcription subunit-like protein(MED9)
		AT2G03070	mediator subunit 8(MED8)
		AT4G22745	methyl-CPG-binding domain 1(MBD1)
		AT1G15340	methyl-CPG-binding domain 10(MBD10)
		AT3G15790	methyl-CPG-binding domain 11(MBD11)
		AT3G63030	methyl-CPG-binding domain 4(MBD4)

		AT5G35330	methyl-CPG-binding domain protein 02(MBD02)
		AT3G46580	methyl-CPG-binding domain protein 5(MBD5)
		AT2G42680	multiprotein bridging factor 1A(MBF1A)
		AT4G16250	phytochrome D(PHYD)
		AT4G18130	phytochrome E(PHYE)
		AT4G36650	plant-specific TFIIB-related protein(PBRP)
		AT1G61040	plus-3 domain-containing protein(VIP5)
		AT1G15200	protein-protein interaction regulator family protein(AT1G15200)
		AT5G02810	pseudo-response regulator 7(PRR7)
		AT2G32080	purin-rich alpha 1(PUR ALPHA-1)
		AT3G12280	retinoblastoma-related 1(RBR1)
		AT4G04920	sensitive to freezing 6(SFR6)
		AT3G17100	sequence-specific DNA binding transcription factor(AT3G17100)
		AT2G33610	switch subunit 3(SWI3B)
		AT2G38560	transcript elongation factor IIS(TFIIS)
		AT2G41630	transcription factor IIB(TFIIB)
		AT4G12610	transcription initiation factor IIF subunit alpha RAP74(RAP74)
		AT5G18230	transcription regulator NOT2/NOT3/NOT5 family protein(AT5G18230)
		AT3G17590	transcription regulatory protein SNF5, putative (BSH)(BSH)
		AT4G10920	transcriptional coactivator p15 (PC4) family protein (KELP)(KELP)
		AT4G32551	transcriptional corepressor LEUNIG(LUG)
		AT3G02860	zinc ion binding protein(AT3G02860)

Supplementary table 5. qPCR primers list.

Gene	Accession	Left primer	Right primer
TSA1	AT3G54640	GGTTCAGTCGCTCTTGAAGG	CCCATCCAGCTATCTGTTTCA
PAL1	AT2G37040	CGCACTTCAGAAGGAACTTATTAGA	ATCGGATACCGGAAAATCCT
PP2A	AT1G13320	GACCGGAGCCAACTAGGAC	AAAACTTGGTAACTTTTCCAGCA
UBC21	AT5G25760	CAGTCTGTGTGTAGAGCTATCATAGCAT	AGAAGATTCCCTGAGTCGCAGTT

Supplementary table 6. Proteins absent in nucleus under control conditions (water treated) and present in nucleus after flg22 treatment

Accession	Description
AT1G07750	RmlC-like cupins superfamily protein
AT1G60670	Protein of unknown function (DUF3755)
AT2G17740	Cysteine/Histidine-rich C1 domain family protein
AT2G23610	ATMES3, MES3 | methyl esterase 3
AT2G44380	Cysteine/Histidine-rich C1 domain family protein

AT3G07720	Galactose oxidase/kelch repeat superfamily protein
AT4G17500	ATERF-1, ERF-1 | ethylene responsive element binding factor 1
AT4G24370	unknown protein
AT5G23060	CaS | calcium sensing receptor
AT5G46780	VQ motif-containing protein
AT5G48657	defense protein-related
AT5G64750	ABR1 | Integrase-type DNA-binding superfamily protein
AT1G06760	winged-helix DNA-binding transcription factor family protein
AT1G09230	RNA-binding (RRM/RBD/RNP motifs) family protein
AT1G13360	unknown protein
AT1G20330	SMT2, CVP1, FRL1 | sterol methyltransferase 2
AT3G13470	TCP-1/cpn60 chaperonin family protein
AT1G64550	ATGCN3, GCN3 | general control non-repressible 3
AT1G69750	COX19-2, ATCOX19-2 | cytochrome c oxidase 19-2
AT1G68340	Protein of unknown function (DUF1639)
AT1G71080	RNA polymerase II transcription elongation factor
AT1G71697	ATCK1, CK, CK1 | choline kinase 1
AT1G79690	atnudt3, NUDT3 | nudix hydrolase homolog 3
AT2G19390	unknown protein
AT2G45880	BMY4, BAM7 | beta-amylase 7
AT3G08530	Clathrin, heavy chain
AT3G12200	AtNek7, Nek7 | NIMA-related kinase 7
AT3G13350	HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain
AT3G17609	HYH | HY5-homolog
AT3G19080	SWIB complex BAF60b domain-containing protein
AT3G21430	ATALY3, ALY3 | DNA binding
AT3G22270	Topoisomerase II-associated protein PAT1
AT3G23280	XBAT35 | XB3 ortholog 5 in Arabidopsis thaliana
AT3G25760	AOC1, ERD12 | allene oxide cyclase 1
AT3G49600	UBP26, SUP32, ATUBP26 | ubiquitin-specific protease 26
AT3G55730	MYB109, AtMYB109 | myb domain protein 109
AT4G18430	AtRABA1e, RABA1e | RAB GTPase homolog A1E
AT4G29230	anac075, NAC075 | NAC domain containing protein 75
AT4G33080	AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein
AT4G34660	SH3 domain-containing protein
AT4G37090	unknown protein
AT5G11430	SPOC domain / Transcription elongation factor S-II protein
AT5G11850	Protein kinase superfamily protein
AT5G16150	GLT1, PGLCT | plastidic GLC translocator
AT5G16300	Vps51/Vps67 family (components of vesicular transport) protein
AT5G20600	FUNCTIONS IN: molecular_function unknown

AT5G22530	unknown protein
AT5G41880	POLA3, POLA4	DNA primases;DNA primases
AT5G42190	ASK2, SKP1B	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 family protein
AT5G47090	unknown protein
AT5G55100	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein

Supplementary table 7. Proteins absent in nucleus under control conditions (water treated), present in nucleus after flg22 treatment and present in nucleus under flg22 +cycloheximide treatments.

Accession	Description
AT1G09230	RNA-binding (RRM/RBD/RNP motifs) family protein
AT3G13470	TCP-1/cpn60 chaperonin family protein
AT1G71080	RNA polymerase II transcription elongation factor
AT1G71697	ATCK1, CK, CK1	choline kinase 1
AT1G79690	atnudt3, NUDT3	nudix hydrolase homolog 3
AT2G45880	BMY4, BAM7	beta-amylase 7
AT3G07720	Galactose oxidase/kelch repeat superfamily protein
AT3G08530	Clathrin, heavy chain
AT3G12200	AtNek7, Nek7	NIMA-related kinase 7
AT3G19080	SWIB complex BAF60b domain-containing protein
AT4G34660	SH3 domain-containing protein
AT5G16300	Vps51/Vps67 family (components of vesicular transport) protein
AT5G23060	CaS	calcium sensing receptor
AT5G42190	ASK2, SKP1B	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 family protein
AT1G60670	Protein of unknown function (DUF3755)
AT2G19390	unknown protein
AT3G21430	ATALY3, ALY3	DNA binding
AT4G24370	unknown protein
AT4G29230	anac075, NAC075	NAC domain containing protein 75
AT4G37090	unknown protein
AT5G11430	SPOC domain / Transcription elongation factor S-II protein

Supplementary table 8. Proteins absent in nucleus under control conditions (water treated), present in nucleus after cycloheximide treatment.

Accession	Description
AT1G52400	BGL1, BGLU18, ATBG1	beta glucosidase 18
AT1G65960	GAD2	glutamate decarboxylase 2
AT1G67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein
AT1G67440	emb1688	Minichromosome maintenance (MCM2/3/5) family protein

AT1G79690	atnudt3, NUDT3 | nudix hydrolase homolog 3
AT2G19390	unknown protein
AT5G45775	Ribosomal L5P family protein
AT2G43500	Plant regulator RWP-RK family protein
AT2G45880	BMY4, BAM7 | beta-amylase 7
AT3G08530	Clathrin, heavy chain
AT3G12140	Emsy N Terminus (ENT)/ plant Tudor-like domains-containing protein
AT3G19080	SWIB complex BAF60b domain-containing protein
AT3G47700	MAG2 | RINT-1 / TIP-1 family
AT3G52960	Thioredoxin superfamily protein
AT5G16300	Vps51/Vps67 family (components of vesicular transport) protein
AT5G19760	Mitochondrial substrate carrier family protein
AT5G23060	CaS | calcium sensing receptor
AT5G41880	POLA3, POLA4 | DNA primases;DNA primases
AT5G42190	ASK2, SKP1B | E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 family protein
ATCG00470	ATPE | ATP synthase epsilon chain
AT1G06530	Tropomyosin-related
AT1G07750	RmlC-like cupins superfamily protein
AT1G09230	U11/U12-65K ribonucleoprotein
AT1G10760	SEX1, SOP1, SOP, GWD1, GWD | Pyruvate phosphate dikinase, PEP/pyruvate binding domain
AT1G15940	Tudor/PWWP/MBT superfamily protein
AT1G24460	unknown protein
AT1G47128	RD21, RD21A | Granulin repeat cysteine protease family protein
AT1G07480	Transcription factor IIA, alpha/beta subunit
AT1G52690	Late embryogenesis abundant protein (LEA) family protein
AT1G54060	ASIL1 | 6B-interacting protein 1-like 1
AT3G13930	Dihydrolipoamide acetyltransferase, long form protein
AT3G13470	TCP-1/cpn60 chaperonin family protein
AT1G56600	AtGolS2, GolS2 | galactinol synthase 2
AT1G60670	Protein of unknown function (DUF3755)
AT1G64110	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G69750	COX19-2, ATCOX19-2 | cytochrome c oxidase 19-2
AT1G68910	WIT2 | WPP domain-interacting protein 2
AT1G69060	Chaperone DnaJ-domain superfamily protein
AT1G71080	RNA polymerase II transcription elongation factor
AT1G71697	ATCK1, CK, CK1 | choline kinase 1
AT1G76040	CPK29 | calcium-dependent protein kinase 29
AT1G76380	DNA-binding bromodomain-containing protein
AT1G77840	Translation initiation factor IF2/IF5
AT1G78020	Protein of unknown function (DUF581)
AT1G80420	ATXRCC1 | BRCT domain-containing DNA repair protein

AT2G03070	MED8 | mediator subunit 8
AT2G12550	| ubiquitin-associated (UBA)/TS-N domain-containing protein
AT2G28450	zinc finger (CCCH-type) family protein
AT2G35330	RING/U-box superfamily protein
AT2G40290	Eukaryotic translation initiation factor 2 subunit 1
AT3G56340	Ribosomal protein S26e family protein
AT2G41350	unknown protein
AT2G42700	FUNCTIONS IN: molecular_function unknown;
AT2G47620	ATSWI3A, CHB1, SWI3A | SWITCH/sucrose nonfermenting 3A
AT3G03980	NAD(P)-binding Rossmann-fold superfamily protein
AT3G13350	HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain
AT3G13990	Kinase-related protein of unknown function (DUF1296)
AT3G19580	AZF2, ZF2 | zinc-finger protein 2
AT3G19860	bHLH121 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT3G21430	ATALY3, ALY3 | DNA binding
AT3G45630	RNA binding (RRM/RBD/RNP motifs) family protein
AT3G55730	MYB109, AtMYB109 | myb domain protein 109
AT3G61140	FUS6, ATFUS6, CSN1, COP11, EMB78, ATSK31, SK31 | 26S proteasome, regulatory subunit Rpn7;Proteasome component (PCI) domain
AT4G02560	LD | Homeodomain-like superfamily protein
AT4G12080	AHL1, ATAHL1 | AT-hook motif nuclear-localized protein 1
AT4G13640	UNE16 | Homeodomain-like superfamily protein
AT4G14710	ATARD2 | RmlC-like cupins superfamily protein
AT4G24370	unknown protein
AT4G24690	ubiquitin-associated (UBA)/TS-N domain-containing protein / octicosapeptide/Phox/Bemp1 (PB1) domain-containing protein
AT4G25880	APUM6, PUM6 | pumilio 6
AT4G27440	PORB | protochlorophyllide oxidoreductase B
AT5G64420	DNA polymerase V family
AT4G29230	anac075, NAC075 | NAC domain containing protein 75
AT4G32160	Phox (PX) domain-containing protein
AT4G36760	ATAPP1, APP1 | aminopeptidase P1 |
AT4G38630	RPN10, MCB1, ATMCB1, MBP1 | regulatory particle non-ATPase 10
AT5G01270	CPL2, ATCPL2 | carboxyl-terminal domain (ctd) phosphatase-like 2
AT5G04200	AtMC9, MC9 | metacaspase 9
AT5G08230	Tudor/PWWP/MBT domain-containing protein |
AT5G11430	SPOC domain / Transcription elongation factor S-II protein |
AT5G13130	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein
AT5G18940	Mo25 family protein
AT5G20600	FUNCTIONS IN: molecular_function unknown
AT5G24890	unknown protein; FUNCTIONS IN: molecular_function unknownstages; BEST Arabidopsis th
AT5G35520	MIS12, ATMIS12 | minichromosome instability 12 (mis12)-like |
AT5G41100	FUNCTIONS IN: molecular_function unknown

AT5G46780	VQ motif-containing protein
AT5G47090	| unknown protein;
AT5G52660	| Homeodomain-like superfamily protein
AT5G61040	unknown protein;

**Supplementary table 9. List of proteins with significant change in abundance between flg22 and control conditions**

| (ANOVA), FDR: 0.05			
F_W (increased abundance after flg22)			
	C: ANOVA Significant	N: -Log ANOVA p value	N: ANOVA q-value
AT1G17210	+	1.95043	0.0494483
AT1G69800	+	2.0884	0.0435278
AT1G80840	+	5.37005	0.02
AT2G17740	+	5.19176	0.0106667
AT2G19390	+	2.80801	0.0204118
AT2G35830	+	2.01496	0.0472051
AT2G37970	+	4.0779	0.00885714
AT2G40140	+	2.88654	0.0189508
AT2G44370	+	3.9951	0.0101333
AT2G44380	+	3.9625	0.00964706
AT3G01830	+	3.48215	0.011
AT5G02790	+	3.78818	0.00990476
AT5G15130	+	3.4282	0.011
AT5G23060	+	2.97237	0.0179636
AT5G25260	+	2.9824	0.0181481
AT5G53000	+	2.19034	0.0374737
AT5G62000	+	4.9921	0.009
W_F (decreased abundance after flg22)			
	C: ANOVA Significant	N: -Log ANOVA p value	N: ANOVA q-value
AT1G75100	+	2.03137	0.0468947
AT2G25450	+	3.10679	0.0143529
AT2G43910	+	4.13671	0.00966667
AT2G45820	+	2.20922	0.0369922
AT3G12390	+	2.07343	0.044274
AT3G45190	+	3.12384	0.01432
AT4G16330	+	2.73554	0.0212105
AT4G22150	+	2.12688	0.0404539
AT4G27900	+	2.13497	0.0401439

Supplementary table 10. List of 57 proteins with significant increase in abundance after flg22 (F) against water (W) and are not significantly increased in abundance in flg22+cycloheximide (FC) against water (W)

Multiple test (ANOVA), FDR: 0.05			
	ANOVA Significant	-Log ANOVA p value	ANOVA q-value
AT1G03210	+	2.45325	0.0369855
AT1G12840	+	2.22348	0.0342143
AT1G18070	+	1.76943	0.0452
AT1G34420	+	2.42251	0.0385556
AT1G41830	+	1.65386	0.0415765
AT1G50480	+	2.06178	0.0371944
AT1G52600	+	2.00287	0.0359059
AT1G59870	+	1.98359	0.0363182
AT1G67360	+	1.63691	0.041906
AT1G76140	+	2.18625	0.0347692
AT2G01470	+	3.27147	0.0292
AT2G05120	+	2.31829	0.0358652
AT2G19190	+	3.1664	0.030963
AT2G21390	+	1.7577	0.0444677
AT2G26780	+	2.31238	0.0348817
AT2G38390	+	1.87317	0.0401165
AT2G39518	+	3.41141	0.024
AT2G43410	+	2.2611	0.0339806
AT2G45540	+	1.59491	0.0422594
AT3G01510	+	1.69746	0.0424857
AT3G03470	+	1.93081	0.0382105
AT3G03610	+	2.36317	0.0365854
AT3G06510	+	2.61263	0.0449565
AT3G21790	+	2.06043	0.0374069
AT3G24503	+	1.36034	0.0468223
AT3G28220	+	2.01109	0.0370123
AT3G29360	+	1.50171	0.0418841
AT3G46280	+	4.49415	0.00933333
AT3G48320	+	3.22733	0.0304348
AT3G51920	+	2.65302	0.0431818
AT4G01290	+	2.50403	0.0393443
AT4G01810	+	2.48301	0.0381818
AT4G02340	+	1.40249	0.0468858
AT4G08770	+	2.51178	0.0397333
AT4G22470	+	2.49955	0.0387302
AT4G24820	+	1.4465	0.0455614

AT4G25970	+	2.60548	0.0436667
AT4G27560	+	1.58015	0.0419048
AT4G28220	+	1.87014	0.0403092
AT4G33090	+	1.51311	0.0422673
AT5G07300	+	1.29476	0.0498378
AT5G11040	+	2.57292	0.0413818
AT5G14930	+	2.31785	0.0354667
AT5G16570	+	1.59414	0.0421379
AT5G20960	+	2.19004	0.0342759
AT5G25260	+	1.80431	0.0441057
AT5G37600	+	1.60581	0.0423176
AT5G39080	+	1.88418	0.0407236
AT5G40760	+	1.54314	0.0418639
AT5G55050	+	6.2199	0
AT5G56730	+	2.25874	0.0343462
AT5G57220	+	3.82078	0.0163333
AT5G60620	+	1.71568	0.0427159
AT5G61900	+	2.01067	0.0363394
AT5G62670	+	3.8612	0.0167273
AT5G64120	+	3.17709	0.0309231
ATCG00720	+	2.162	0.035328
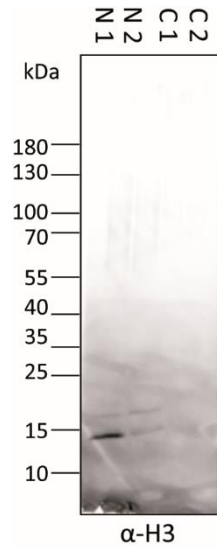
Supplementary figure 1. Non-cropped western blot (of Figure 10 B) of nuclear and cellular proteins with anti-Histone H3 antibody. Two independent experiments are shown, N denotes nuclear and C cellular.
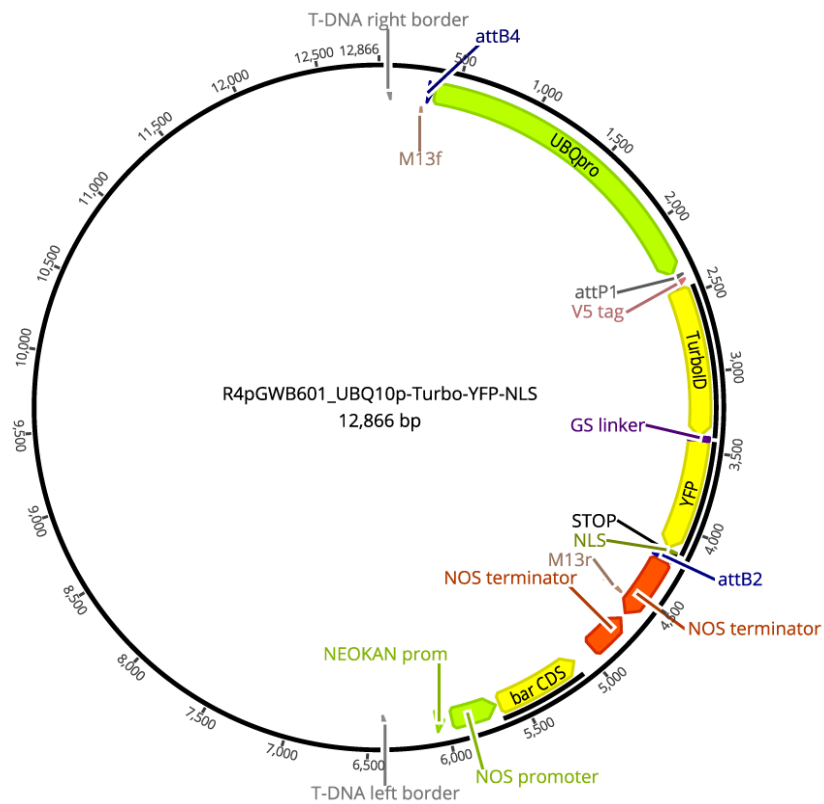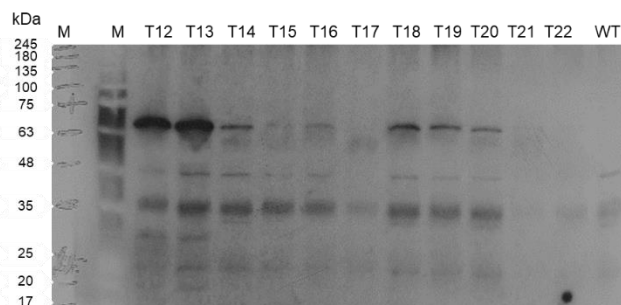

Supplementary figure 2. Vector for expressing TurboID-YFP-NLS in plants under UBQ10 promoter. (Addgene: 127368).
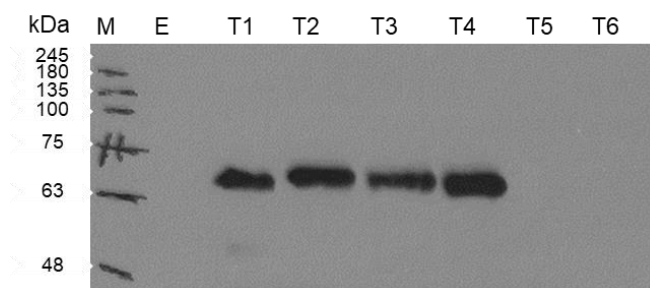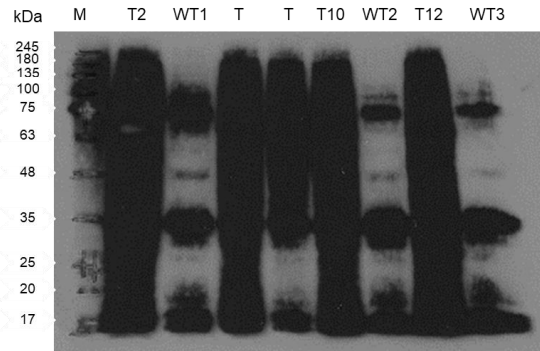
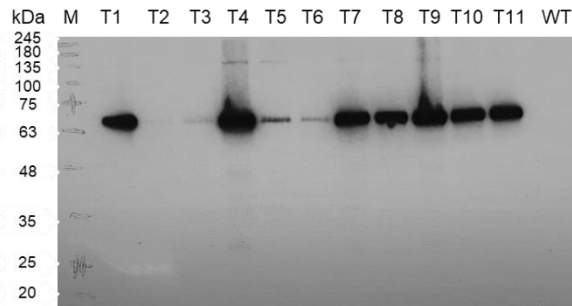**Supplementary figure 3.** Non-cropped western blot (of Figure 14).
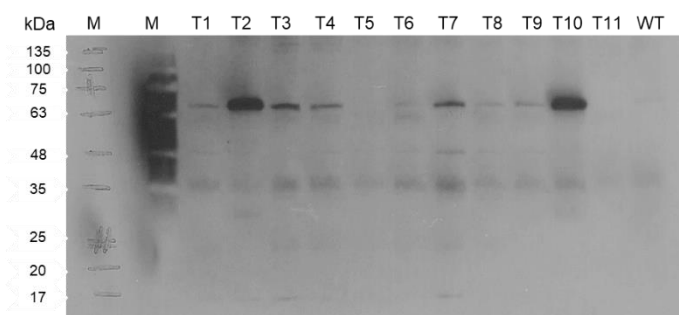
Supplementary figure 4. Non-cropped western blot (of Figure 17 B, left).
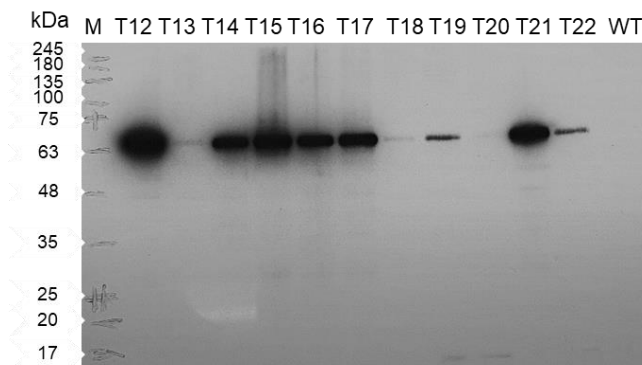

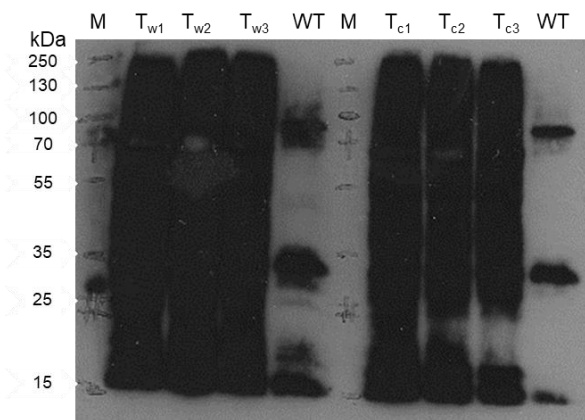**Supplementary figure 5.** Non-cropped western blot (of Figure 17 B, right).

Supplementary figure 6. Non-cropped western blot (of Figure 17 C).

Supplementary figure 7. Non-cropped western blot (of Figure 21 B, left).

Supplementary figure 8. Non-cropped western blot (of Figure 21 B, right).

Supplementary figure 9. Non-cropped western blot (of Figure 21 C).

Supplementary figure 10. Protein markers used. Left: Ladder blue eye prestained protein marker (Jena biosciences, PS-104). Right: Page ruler plus, prestained protein ladder (Thermo fisher, 26619).

152

Acknowledgment

I would like express my deepest thanks to my direct supervisor and bigger brother Dr. Wolfgang Hoehenwarter for his support and guidance throughout the years, his door was always open whenever I had questions about my research or writing, his comments and remarks had a huge impact throughout the whole period of this work.

I would like to express my deep gratitude to my university supervisor Prof. Dr. Ingo Heilmann for the useful comments, remarks and supervision and for his guidance generally as the speaker of RTG 2498 and as a member of my Thesis Committee.

I would like also to express my gratitude to Prof. Dr. Kristina Kühn for her comments and remarks during my Thesis committee meetings. Additionally, I must thank Prof. Dr. Tina Romeis for allowing me to continue my research project in the BPI department.

I would like to give special thanks to my friend and lab colleague Mohammad Abukhalaf for the fruitfull discussions and for our collaboration. Also, I would like to thank Domenika Thieme, Carsten Proksch, Nicole Bauer and Sylvia Krüger for their help and support in the lab work and LC-MS analysis.

I would like to thank Prof. Dr. Bettina Hause and Hagen Stellmach for introducing me to the microscopy devices. Special thanks to Dr. Justin Lee for the valuable discussions.

I would like to acknowledge Dr. Martin Schattat and Dr. Marieke Heilmann for their useful comments and discussions and for their contribution in the published manuscript. I am also grateful to my co-workers at IPB for providing advice, comments and a professional working environment.

Finally, I must express my very profound gratitude to my parents, brother, friends and my love Luise Reichert for providing me with support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

<div align="center">**CV**</div>

MOHAMED ADEL ABDELAZIZ AYASH

Date and place of Birth: 19.08.1985. Cairo, Egypt.
Nationality: Egyptian

WORK EXPERIENCE

- **Research Assistant (PhD), Proteome Analytics**
  Leibniz Institute for Plant Biochemistry, Halle (Saale)
  Martin Luther University of Halle Wittenberg (RTG 2498)
  Aug 2019 – Jan 2023

- **QC specialist**
  Sanofi, Egypt
  Jul 2014 – Sep 2015

- **R&D specialist**
  Sigmatec Pharmaceutical industries, Egypt
  Apr 2009 - Jul 2014

- **Production pharmacist**
  El-Nile Co For Pharmaceuticals And Chemical Industries, Egypt
  Jan 2009 - Apr 2009

- **R&D specialist**
  Mepaco pharmaceuticals
  Dec 2007 - May 2008

EDUCATION

- **Martin Luther University Halle-Wittenberg**
  Fraunhofer IZI
  Oct 2015- Dec 2018
  Masters of Pharmaceutical and Industrial Biotechnology
  Master Thesis " Cloning and heterologous expression of human Ovastacin in *E. coli*,
  Pichia Pastoris and S2 cells"

- **Cairo University, Faculty of Pharmacy**
  Bachelor's degree in Pharmaceutical sciences
  Sep 2002- May 2007

PRESENTATIONS & PUBLICATIONS

- **AYASH, M**., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. 2021. LC-MS Based Draft Map of the *Arabidopsis thaliana* Nuclear Proteome and Protein Import in Pattern Triggered Immunity. Frontiers in plant science, 12, 744103-744103.

- Abukhalaf, M., Bassal, M., Majovsky, P., Thieme, D., Herr, T., **Ayash, M**., Tabassum, N., Al Shweiki, M. R., Proksch, C., Hmedat, A., Ziegler, J., Lee, J., Neumann, S. & Hoehenwarter, W. 2020. Reshaping of the Arabidopsis thaliana Proteome Landscape and Co-regulation of Proteins in Development and Immunity. Mol Plant, 13, 1709-1732. doi: 10.1016/j.molp.2020.09.024

- Poster presentation in Proteomic sciences, Annual Congress of the European Proteomics Association (EuPA), Germany (April 2022)

Statement of authorship and good scientific practice

I declare that the PhD thesis submitted is my own work and was done independently. I did not use other than the listed support / resources. All sources and support used are acknowledged as references or quoted in the acknowledgement.

With this dissertation I am applying for the first time to obtain a doctoral degree.

_____

Signature, Location and Date

Dr. Marieke Heilmann performed one western blot (Figure 10 B).

Domenika Thieme and Carsten Proksch were the technical assistance for the MS (calibration and maintenance).

Dr. Martin Schattat provided me with a sorted list of the *Arabidopsis thaliana* genes which were analyzed by the LOCALIZER tool.

This PhD thesis including all parts was written by myself.

The published manuscript in Frontiers in plant Science was written by myself except for the introduction.

Dr. Wolfgang Hoehenwarter supervised the project.

Parts of this thesis was published in the following journal and was presented as a poster in a conference.

Paper

AYASH, M., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. 2021. LC-MS Based Draft Map of the Arabidopsis thaliana Nuclear Proteome and Protein Import in Pattern Triggered Immunity. Frontiers in plant science, 12, 744103-744103. Licensed under CC BY 4.0.

Poster

AYASH, M., ABUKHALAF, M., THIEME, D., PROKSCH, C., HEILMANN, M., SCHATTAT, M. H. & HOEHENWARTER, W. LC-MS Based Draft Map of the Arabidopsis thaliana Nuclear Proteome and Protein Import in Pattern Triggered Immunity. XIV Annual Congress of the European Proteomic Association (EuPA), 3 - 7 April 2022, Germany.