Journal

Plant Biotechnology Journal (2013) 11, pp. 1069–1079 doi: 10.1111/pbi.12101

Zinc finger artificial transcription factor–based nearest inactive analogue/nearest active analogue strategy used for the identification of plant genes controlling homologous recombination

Qi Jia[1,†], Marcel C. van Verk[2,†], Johan E. Pinas, Beatrice I. Lindhout[3], Paul J.J. Hooykaas and Bert J. van der Zaal*

Department of Molecular and Developmental Genetics, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

Received 7 May 2013;
accepted 27 June 2013.
**Correspondence* (Tel +31715274760;
fax +31715274999;
email b.j.v.d.zaal@biology.leidenuniv.nl)
[1]Present address: College of Crop Science, Fujian Agriculture and Forestry University, Shangxiadian Road 15, 350002, Fuzhou, China.
[2]Present address: Plant-Microbe Interactions, Department of Biology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands
[3]Present address: Rijk Zwaan Breeding B.V., Eerste Kruisweg 9, 4793RS, Fijnaart, The Netherlands
[†]These authors contributed equally to this work.

Keywords: somatic intrachromosomal homologous recombination, zinc finger artificial transcription factor, genome interrogation, mutagenesis, transcriptome analysis, next-generation sequencing.

Summary

In previous work, we selected a particular transcription factor, designated VP16-HRU, from a pool of zinc finger artificial transcription factors (ZF-ATFs) used for genome interrogation. When expressed in *Arabidopsis thaliana* under control of the ribosomal protein S5A promoter, the *RPS5A::VP16-HRU* construct led to a 200- to 300-fold increase in the frequency of somatic intrachromosomal homologous recombination (iHR). Because the expression of each ZF-ATF leads to a large number of transcriptional changes, we designed a strategy employing a collection of structurally similar ZF-ATFs to filter out the transcriptional changes relevant to the phenotype by deep sequencing. In that manner, 30 transcripts were found to be consistently induced in plants with enhanced homologous recombination (HR). For 25 of the cognate genes, their effect on the HR process was assessed using cDNA/gDNA expression constructs. For three genes, ectopic expression indeed led to enhanced iHR frequencies, albeit much lower than the frequency observed when a HR-inducing ZF-ATF was present. Altogether, our data demonstrate that despite the large number of transcriptional changes brought about by individual ZF-ATFs, causal changes can be identified. In our case, the picture emerged that a natural regulatory switch for iHR does not exist but that ZF-ATFs-like VP16-HRU act as an ectopic master switch, orchestrating the timely expression of a set of plant genes that each by themselves only have modest effects, but when acting together support an extremely high iHR frequency.

Introduction

The process of homologous recombination (HR) is comprised of mechanisms that allow and facilitate molecular exchanges of genetic material based upon the presence of identical or nearly identical DNA sequences. HR has been described in virtually all types of living organisms and is commonly regarded as a potent force that shapes all kinds of genetic elements, including the layout of complex genomes. From a biotechnological point of view, precise HR-mediated integration of a DNA sequence of interest at a particular site within a genome is the ultimate tool for genetic engineering. However, in many higher eukaryotic organisms and particularly in plants, virtually all incoming DNA sequences integrate at random positions within the genome rather than being directed by sequence homology.

More and more evidence showed that the genes and gene products that are mechanistically involved in the HR process itself are highly conserved throughout eukaryotes (Bleuyard *et al.*, 2006; Lieberman-Lazarovich and Levy, 2011; Markmann-Mulisch *et al.*, 2007; Shaked *et al.*, 2005). Therefore, it seems likely that differences in HR frequencies must be caused by differences in the regulation of the process rather than by the lack of particular components. Unfortunately, true endogenous master regulators of HR events—if they exist—are completely unknown, thus hampering the understanding as well as the application of HR.

In the last few years, zinc finger (ZF) technology has gained enormous attention within the field of functional genomics (Lee *et al.*, 2010; Sera, 2009; Shukla *et al.*, 2009). The technology is based upon the availability of well-characterized 30 amino acid ZF moieties, each with a cognate three-base pair DNA-binding site, combined with the possibility to generate multifingered (poly-dactyl) PZF domains by fusing individual ZF domains (Dreier *et al.*, 2001, 2005; Pavletich and Pabo, 1991; Segal *et al.*, 1999). The length of the DNA sequence recognized thus depends upon the numbers of ZFs fused. With PZFs available, an extremely useful toolbox for molecular genetics has been created, ranging from the ability to create site-specific ZF nucleases (ZFNs) to the possibility to generate ZF-based artificial transcription factors (ZF-ATFs) by fusing transcriptional activator or repressor domains with a PZF domain (http://www.zincfingers.org) (Wright *et al.*, 2006).

An intriguing possibility to employ ZF-ATFs in genomic research is the introduction of libraries of different ZF-ATFs into an organism of interest. For three-fingered (3F) ZF-ATFs, each ZF-ATF potentially interacts with the multitude of cognate 9-bp sites within a eukaryotic genome. Hence, supposing that one is able to generate a collection of cells or organisms each expressing a different 3F ZF-ATF, the use of even a relatively simple collection of 3F ZF-ATFs more or less ensures that all available genomic sites where an interaction can result in altered expression of a nearby gene are targeted. Introducing such a library into host cells has therefore been called 'genome interrogation', a kind of brute force approach to reveal hidden phenotypic possibilities as novel dominant traits (reviewed in Beltran *et al.*, 2006). Once a desired phenotype is selected, the causal ZF-ATF can easily be identified, thereby bringing the desired novel trait under experimental control. We demonstrated previously that this approach is also applicable at the level of a multicellular eukaryotic organism, in this case the model plant *Arabidopsis thaliana*. In that study (Lindhout *et al.*, 2006), a ZF-ATF was identified capable of increasing somatic intrachromosomal homologous recombination (iHR) by on average 200- to 300-fold when assayed in an HR indicator line containing an integrated interrupted *GUS* reporter gene. To our knowledge, this large increase in iHR frequency has by far not been equalled by any other iHR mutant reported thus far. The mutant seedlings were more resistant to the DNA double-strand break (DSB) inducing compound bleomycin and were quite normal regarding their overall phenotype. The specific ZF-ATF found to be responsible for the enormously enhanced iHR phenotype was designated VP16-HRU, referring to the HR-up-regulated phenotype and to its VP16 transcriptional activation domain (Lindhout *et al.*, 2006). By studying VP16-HRU-expressing plants, it should be possible to learn a lot regarding the iHR mechanism.

Genome interrogation thus allows for the discovery of unique phenotypes of interest in an unbiased manner, without any a priori ideas regarding the required endogenous changes responsible for the phenotype. However, finding the mechanism by which a particular PZF-ATF leads to a novel phenotype is extremely challenging. In *Arabidopsis*, a typical 3F PZF domain as present in VP16-HRU might already find about 1000 loci that contain a cognate 9-bp recognition sequence (5′GTGGAGGCT in this case) and can thus be expected to cause a multitude of changes in any of the 'omics'. In preliminary transcriptome analyses using microarrays, we indeed found large numbers of genes to be differentially expressed due to VP16-HRU expression (Lindhout, 2008).

When having access to different mutants sharing a similar phenotype after genome interrogation, looking for shared transcriptional changes can provide a focus towards getting to know the possibly causal changes leading to this phenotype (Lee *et al.*, 2011). Unfortunately, when having just a single mutant identified, further analysis of relevant ZF-ATF-induced transcriptional changes is impossible. In the present study, we devised a strategy to overcome such a standstill, inspired by the reported nonexclusivity of ZF–DNA interactions (Carroll *et al.*, 2006; Segal *et al.*, 1999). Lack of complete specificity should allow for a dedicated selection of highly similar ZF-ATFs, which are still able to induce the phenotype of interest [nearest active analogues (NANs)] or not [nearest inactive analogues (NINs)]. In this study, we provide proof of principle for such a 'NIN/NAN strategy'. Apart from the original VP16-HRU, we found three analogues also inducing iHR. Comparing the transcriptomes of plants expressing the four NANs to those of plants expressing NINs, we were able to identify endogenous plant genes involved in iHR.

Results

Generating structural analogues of VP16-HRU and assessing iHR frequency

When PZF modules are built using our system (Neuteboom *et al.*, 2006), each ZF domain is added as a C-terminal extension to the previous one(s) cloned. For convenience, each ZF moiety can be given a number referring to its identity (Segal *et al.*, 1999); the 3F PZF domain that is present in VP16-HRU is in our system generated by sequential addition of ZF-encoding oligonucleotides 9 (recognizing 'GTG'), 5 (recognizing 'GAG') and 15 (recognizing 'GCT'), resulting in the domain 9-5-15.

Considering reported data (Carroll *et al.*, 2006; http://www.zincfingers.org/), 41 ZFs were available with sufficient affinity and specificity: 16 for GNN, 12 for ANN, 2 for TNN and 11 for CNN. Therefore, starting from the 2F sequence 9-5, a 9-5-X series of 3F containing HRU derivatives was formed, by using for 'X' in total 41 double-stranded oligonucleotide primers encoding these particular ZFs. For unknown reasons, the combination 9-5-3 could not be assembled in the high-copy cloning vector pSKN-SgrAI (Neuteboom *et al.*, 2006). Hence, 40 constructs encoding 3F modules were generated that were subsequently inserted as *Sfi*I fragments into the T-DNA-containing binary plasmid vector pRF-VP16 (Lindhout *et al.*, 2006). Additionally, we generated constructs with the 2F modules 9-5 and 5-15. All constructs were subsequently introduced into *Arabidopsis* iHR reporter line 1406 (Gherbi *et al.*, 2001) using floral dip transformation.

For each construct, we strived for assessing about 65 individual primary transformed (T1) seedlings for the frequency of iHR events by staining whole seedling for GUS activity. Just as observed previously (Lindhout *et al.*, 2006), iHR events were predominantly found in aerial parts of seedlings, hardly ever in roots.

An overview of the average number of recombination events in seedlings at 9–10 days post germination (DPG) is given in Figure 1. The underlying data are available in Table 1. T1 transformants generated with the original VP16-HRU construct (with PZF 9-5-15) exhibited on average 3.20 recombination events. Three other constructs induced a comparably enhanced rate of recombination and therefore can be classified as NANs:

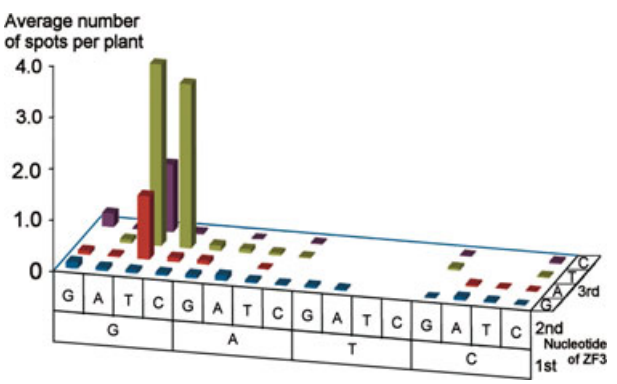

Figure 1 Average number of recombination events (GUS-positive spots) per different construct determined for about 65 primary transformed (T1) seedlings at 9–10 days postgermination (9–10 DPG) per construct. Seedlings were selected from kanamycin-containing selective plates. All constructs share the sequences encoding the first two ZFs, that is, the combination 9-5. The 3rd varying ZF is indicated within the graph by its 1st, 2nd and 3rd optimal recognition nucleotide within the DNA.

Table 1 Recombination events scored in primary transformants for the different VP16-containing ZF-ATFs of the 9-5-X type at 9–10 DPG

3rd ZF	Spots	Plants with spots	Plants	Spots per plant	% Plants with spots	3rd ZF	Spots	Plants with spots	Plants	Spots per plant	% Plants with spots
GGG (1)	4	3	32	0.125	9.4	AAC	0	0	43	0.000	0.0
GGA (2)	2	2	27	0.074	7.4	ATG	2	2	43	0.047	4.7
GGC (4)	25	8	95	0.263	8.4	ATA	1	1	89	0.011	1.1
GAG (5)	2	2	29	0.069	6.9	ATT	2	2	41	0.049	4.9
GAA (6)	2	2	65	0.031	3.1	ACA	1	1	36	0.028	2.8
GAT (7)	1	1	14	0.071	7.1	ACT	1	1	54	0.019	1.9
GAC (8)	0	0	74	0.000	0.0	ACC	1	1	61	0.016	1.6
GTG (9)	7	6	117	0.060	5.1	TGG	2	2	47	0.043	4.3
GTA (10)	47	19	38	1.237	50.0	TAG	2	2	64	0.031	3.1
GTT (11)	99	22	28	3.536	78.6	CGG	0	0	73	0.000	0.0
GTC (12)	154	44	117	1.316	37.6	CGT	2	2	43	0.047	4.7
GCG (13)	5	4	86	0.058	4.7	CGC	1	1	75	0.013	1.3
GCA (14)	8	7	94	0.085	7.4	CAG	4	4	67	0.060	6.0
GCT (15)	336	53	105	3.200	50.5	CAA	4	4	106	0.038	3.8
GCC (16)	4	4	131	0.031	3.1	CTG	3	3	103	0.029	2.9
AGG	8	8	124	0.065	6.5	CTA	1	1	95	0.011	1.1
AGA	5	5	67	0.075	7.5	CCG	0	0	40	0.000	0.0
AGT	4	4	39	0.103	10.3	CCA	0	0	35	0.000	0.0
AAG	4	4	38	0.105	10.5	CCT	1	1	45	0.022	2.2
AAT	3	3	42	0.071	7.1	CCC	2	2	58	0.034	3.4

ZF-ATFs, ZF-based artificial transcription factors; DPG, days after germination.

9-5-10 with 1.24 events, 9-5-11 with 3.54 events and 9-5-12 with 1.32 events. All these constructs contained PZF modules exclusively consisting of GNN-recognizing ZFs.

All 9-5-X PZF domains with ANN-, TNN- or CNN-recognizing PZF modules only yielded ZF-ATF genes leading to iHR frequencies of at most 0.075 spots per seedling (Figure 1 and Table 1). These frequencies were comparable to those observed in primary transformants expressing VP16 without a fused ZF sequence (0.051 spots per seedling) or with fused 9-5 and 5-15 2F domains (0.080 and 0.049 spots per seedling, respectively). For several constructs, not even a single iHR event was observed even in a large number of primary transformants. Apparently, the process of selecting primary transformants did not necessarily induce iHR events. For the parental 1406 line, checked at 10 DPG when growing on medium lacking kanamycin selection, the iHR frequency was at most about 0.01 event per seedling, as found previously (Lindhout *et al.*, 2006). Hence, it can be stated that in the present study, NAN-expressing seedlings exhibited at least a 100-fold higher iHR frequency than controls.

As for the choice of the most closely related ZF-ATFs not leading to enhanced iHR, the NINs, we selected those from the constructs where 'X' was a GNN-recognizing ZF, thus truly the most similar molecular neighbours of the NANs: 9-5-9, 9-5-13, 9-5-14 and 9-5-16 with 0.060, 0.058, 0.085 and 0.031 iHR events per seedling, respectively.

Experimental set-up and transcriptome sequencing

For getting access to plant material representing the different NIN/NAN constructs or control constructs expressing VP16 fusions with just the first two (9-5) or the last two (5-15) ZFs of the 9-5-15 combination, 14–16 independent plant lines were selected for each construct at the T2 stage. Most of these lines possessed the ZF-containing transgene as a single Mendelian trait, segregating 3 : 1 for kanamycin resistance. The overall pattern of iHR frequency at the T2 stage closely resembled the T1 data. For all lines collectively selected to represent particular constructs for transcriptome analysis, T3 seeds of 4-6 kanamycin-resistant T2 plants were pooled. Subsequently, equal amounts of T3 seeds from each of the 14-16 T3 lines were pooled to obtain well-balanced super pools. For convenience, we will further refer to these balanced pools that represent a particular constructs as 'lines'. On average, only one in six seedlings of these lines will lack the dominant ZF-ATF-encoding transgene, a fraction small enough to have little effects upon the outcome of transcriptome analysis. Seedlings can thus be raised without prior antibiotic selection as such a selection will cause an unnecessary bias when comparing transcriptome data with data from the parental HR reporter line 1406 containing only the iHR reporter construct.

To verify that iHR frequencies were still conserved in the T3 generation and were also observable in nonselected soil-grown seedlings, plant material from aerial parts of 20 T3 seedlings was collected at 9 DPG (mostly consisting of cotyledons at this stage) and was GUS-stained for iHR frequency. Just as for the T2 seedlings, the iHR frequencies were similar to those observed in T1 seedlings (Table 2). Hence, we decided to use soil-grown seedlings for RNA extraction. Apart from avoiding kanamycin selection, the growth conditions for seedlings of all lines should then be more similar than when using tissue-cultured material, wherein between-plate differences might contribute to variation. Each pooled T3 line was represented by 20 pots with ten seedlings each, just as the parental iHR reporter line 1406. All pots were put at random positions. Once the seedlings had reached an age of 20 DPG, all green plant parts were harvested for RNA isolation in a fully randomized order to avoid major influences of the circadian rhythm on the transcriptome of different lines.

Two main different methods for next-generation sequencing (NGS) of mRNA exist: RNA-Seq and digital gene expression (DGE).

Table 2 Recombination events in 20 seedlings at 9 DPG

Type	Line	Recombination events	Sum	Average
Control	1406	20 × 0	0	0.00
Control	5-15	20 × 0	0	0.00
Control	9-5	20 × 0	0	0.00
NIN	9-5-9	1 × 1, 19 × 0	1	0.05
NIN	9-5-13	20 × 0	0	0.00
NIN	9-5-14	20 × 0	0	0.00
NIN	9-5-16	20 × 0	0	0.00
NAN	9-5-10	3 × 1, 2 × 2, 1 × 3, 1 × 4, 13 × 0	14	0.70
NAN	9-5-11	2 × 1, 2 × 2, 2 × 3, 1 × 6, 1 × 8, 2 × 10, 1 × 12, 9 × 0	58	2.90
NAN	9-5-12	4 × 1, 1 × 2, 2 × 5, 1 × 6, 1 × 13, 11 × 0	35	1.75
NAN	9-5-15	4 × 1, 3 × 2, 2 × 3, 1 × 4, 10 × 0	20	1.00

DPG, days after germination; NAN, nearest active analogues; NIN, nearest inactive analogues.

With RNA-Seq, the entire mRNA is sequenced, and therefore, the total number of reads obtained is besides its expression level also highly dependent on the length of the transcript. While with DGE, which is based upon added tags for serial analysis of gene expression (SAGE tags), only one read per tag will be obtained for each mRNA, which removes the gene length dependency. We chose to use DGE in order to obtain as much in-depth information as possible.

Transcriptome analysis

As mentioned above, our research plan employing NIN/NAN constructs was aimed to acquire a clear focus on causal changes in gene expression patterns related to iHR. When defining possible causal changes in gene expression, we deliberately wanted to avoid experimental bias due to any subjective criteria for the selection of genes of interest. To fully appreciate the method of genome interrogation in combination with the NIN/NAN strategy, the selection criteria should be simple and should result in a manageable collection of 25–30 genes for which the individual iHR-inducing potential can be assessed by ectopical expression in plants.

With the certainty that NAN-type ZF-ATFs act as ectopic dominant master switches of iHR frequency, the following assumptions can be made: (i) NANs act by directly or indirectly affecting the expression of an endogenous iHR master switch or (ii) an endogenous master switch does not exist, but NANs act by fortuitous differential regulation of multiple genes that together are directly or indirectly involved in iHR.

When (i) is true and supposing that up-regulation of mRNA levels is required for increased iHR, the endogenous master switch should be up-regulated exclusively in NANs. Alternatively, with (ii) being true, collectively iHR-inducing endogenous genes are likely to share a characteristic expression pattern in all NAN lines. However, although largely hypothetical, it might be that the most closely related NIN lines of 9-5-X signature lack iHR induction by improper expression of just one or a few of the otherwise collectively iHR-inducing genes. Subtraction of NIN and control data from NAN data should therefore be executed with care.

As a standard, the minimum expression level that is considered as expressed at significant levels is three transcripts per million sequences (TPM) (Meyers *et al.*, 2004). In our case, where crucial steps in iHR induction might only take place in a small subset of cells where the expression profile is just right and timely, we lowered this to 0.2 TPM before considering a tag as being expressed. With a minimal number of over 25 million tags sequenced, 0.2 TPM still represents a minimum cut-off of five tags in total per sample.

Besides setting the minimum expression level to 0.2 TPM for each of the NAN lines, some additional but still relatively mild criteria were used to identify possible iHR-inducing genes; in each of the NAN lines, transcripts corresponding to tags should at least be 2.0-fold up- or down-regulated compared with the levels in the parental iHR reporter line and at least 1.75-fold up- or down-regulated compared with the 2F control line 9-5. The analysis of the data was further performed on two separate levels. Firstly, all expression levels from tags that uniquely correspond to a single ORF were summarized into a single expression value per ORF that was subsequently used for calling differential expression. Secondly, the expression levels of all tags were not summarized based on ORFs, but were further analysed individually on a per tag basis. This may result in some tags originating from the same ORF having different levels of differential expression.

A summarization of all differentially expressed ORFs/tags found in the different NAN lines is given in Figure 2. Based on the summarized ORFs, roughly 600 are differentially up- or down-regulated compared with controls (Figure 2a). Striking is that although the predicted DNA recognition sequence of each of the 3F domains differs only by a single nucleotide, the number of ORFs that are uniquely differentially expressed within each of the lines is on average more than 50%. On the other hand, the number that is shared among the four NAN lines is only around 10%. This reduction for the up-regulated ORFs from around 600 to just 68 shared up-regulated ORFs indicates the strength of the NIN/NAN approach. Although the total numbers for the analysis based on tags are much larger compared with ORFs (~10–20×), they represent a similar number of ORFs and therefore also have a similar complexity and complexity reduction as the summarized ORFs (Figure 2b).

With 68 gene tags corresponding to up-regulated mRNAs shared between the NAN transcriptomes and 26 tags shared for down-regulated mRNAs (Figure 2a), the cognate gene list was still too large for pursuing individual tests regarding iHR potential. Therefore, additional selection criteria were set for both summarized ORF and tag-based differential expression lists. As a SAGE tag should in principle originate from the 3′ end of a transcript, all tags originating from another region of the transcript than the 3′ end were discarded. Furthermore, as none of the NIN lines exhibited an enhanced iHR phenotype at a level comparable to

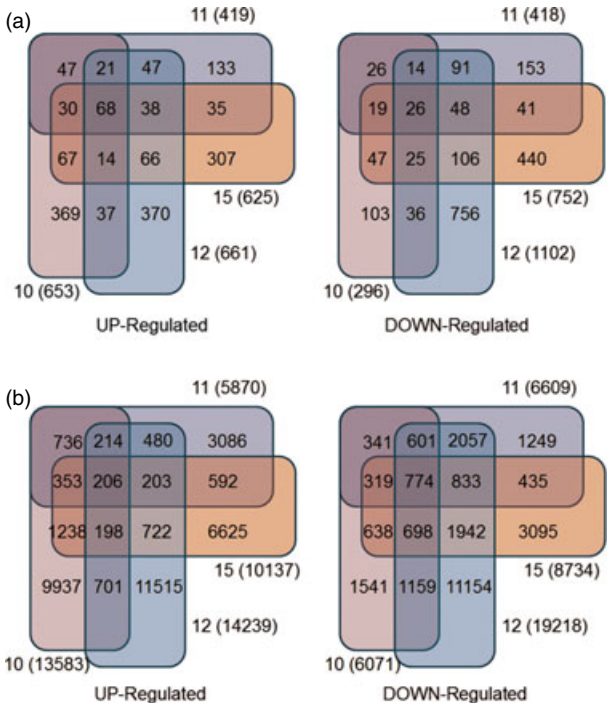

Figure 2 Venn diagram of overlapping differentially expressed genes within the nearest active analogues (NAN) lines. (a) Venn diagram of differentially expressed genes based upon the summarization of tags per ORF. (b) Venn diagram of differentially expressed tags, where multiple tags can represent the same ORF. Numbers outside of the Venn diagram represent the zinc finger NAN line 9-5-(10/11/12/15), and the numbers in parentheses represent the total number of differentially expressed genes or tags respectively for panel (a) and (b). Differentially expressed genes or tags differed at least twofold in expression level from 1406 parental control and at least 1.75-fold with the 2F 9-5 line, with a minimum of 0.2 TPM, as also indicated in text.

the NAN lines, all summarized ORFs and tags that had a lower expression for any of the NAN lines in comparison with any of the NIN lines were discarded. After this additional selection, the remaining up- and down-regulated candidates from both sets were combined. In case of an overlap between the two sets, the expression profile of the summarized ORF was used. An overview of the 30 up- and 14 down-regulated iHR candidates that remained after this additional selection is given in Table 3, with a full description in Table S1.

Further selecting the candidate genes for ectopic expression

When using ZF-ATFs with a VP16 transcriptional activator domain, the primary transcriptional changes are likely to be enhancement of gene expression. Therefore, we did not further pursue analysis of down-regulated genes. As for the 30 up-regulated genes given in Table 3, the targets obtained from only tag data were critically examined. Four of those tags corresponding to the genes At3g44790, At5g01600, At5g07510 and At5g13950 showed extremely low expression levels compared with other tags corresponding to the same gene and therefore were not considered for further experiments. The ORF of At2g24440 (a predicted selenium binding protein) could not be obtained by PCR using either cDNA or genomic DNA as a template.

All remaining 25 ORFs (highlighted in grey in Table 3) were successfully amplified and cloned within the expression cassette of the binary vector. As the cloning strategies resulted in rather high GC-rich regions just upstream of the ATG start codon, an AMV leader sequence was introduced to enhance translation (Datla *et al.*, 1993).

For ectopic expression in *Arabidopsis* plants, ORF representing DNA sequences were put under control of the meristem-specific *RPS5A* promoter (Weijers *et al.*, 2001), just as the original VP16-HRU (Lindhout *et al.*, 2006) as well as all analogues tested above, or under control of the *CaMV35S* promoter for constitutive expression.

Homologous recombination in plants ectopically expressing selected genes

For those 25 ORFs for which plant expression constructs could be made, primary transformants with *RPS5A* (R) or *CaMV35S* (35S) promoter-driven gene constructs were selected on kanamycin-containing medium. Primary transformants resulting from empty vector transformation were selected and grown similarly, but the parental iHR reporter line 1406 was grown on medium lacking kanamycin. After 10 or 20 days of growth, iHR events were visualized by GUS staining and counted (Table 4).

With *Agrobacterium*-mediated transformation causing random insertions of transgenes into the host genome, GUS staining for iHR events at the level of individual primary transformed seedlings can be considered as an independent assay for the ability of particular ORF expression constructs to induce iHR. As can be seen in the Table 4, the number of seedlings representing the constructs was usually fairly large, and most GUS-positive seedlings exhibited only a single recombination event, indicating that the variance of the data for each gene construct is small. ANOVA with LSD could thus be applied to test whether iHR frequency due to transformation with expression constructs differed from controls. Graphs representing the results are given in Figure 3 and Figure S1.

At 10 DPG (Figure S1), the average iHR frequency was low in most primary transformants. In the present data set, only three of fifty constructs led to an iHR frequency that was higher than control values with P-value < 0.01: R-At1g66300, 35S-At2g18193 and 35S-At5g15800. At 20 DPG (Figure 3), the overall iHR frequency was several fold higher. At this stage, eight *RPS5A* promoter-driven expression constructs caused enhanced iHR frequencies with $P < 0.01$: R-At1g55380, R-At1g66300, R-At2g18193, R-At2g22630, R-At3g20360, R-At3g21260, R-At3g28910 and R-At5g48720. Three CaMV35S promoter-driven constructs, 35S-At1g55380, 35S-At1g66300 and 35S-At2g22630, did lead to enhanced iHR frequency as well ($P < 0.01$). The ORFs of the latter constructs, encoding a cysteine/histidine-rich C1 family protein (At1g55380), an F-box/RNI-like/FBD-like domains-containing protein (At1g66300) and transcription factor Agamous-like 17 (At2g22630), were in fact able to evoke iHR events irrespective of the promoter that was used and did so with a P-value < 0.001. The only other construct leading to a similarly significant change was R-At3g21260, encoding glycolipid transfer protein 3, but this occurred concomitant with early ageing and necrosis. Because enhanced iHR in our NAN lines was never found to be related to any clear phenotypic change, certainly not of the kind induced by R-At3g21260, it therefore seems that iHR data for R-At3g21260 have to be interpreted with caution; they might reflect an iHR response to some extreme physiological stress.

Table 3 Overview of selected (validated in grey) candidate genes and their expression in TPM

UP-regulated	ORF-based list	TAG-based list	Control			NIN				NAN			
			1406	9-5	5-15	9-5-9	9-5-13	9-5-14	9-5-16	9-5-10	9-5-11	9-5-12	9-5-15
AT1G11070	√	√	0.93	2.03	1.98	1.92	1.16	1.39	1.45	3.73	6.41	4.58	6.88
AT1G55380	√	√	1.38	1.43	1.58	2.43	1.22	1.26	1.23	4.53	5.47	3.27	5.47
AT1G65300	√	√	0.26	0.13	0.22	0.20	0.35	0.26	0.29	1.36	1.76	1.04	0.66
AT1G66300	√	√	0.08	0.17	0.00	0.12	0.09	0.26	0.00	1.20	1.95	0.84	1.87
AT2G18193		√	0.78	0.77	2.12	1.06	0.72	0.65	0.76	1.73	5.73	4.50	5.97
AT2G22630	√		0.00	0.03	0.07	0.12	0.03	0.03	0.00	0.36	0.41	0.24	0.39
AT2G24440	√	√	0.78	1.44	1.36	1.92	1.94	2.17	2.32	2.93	3.71	6.09	3.70
AT2G27880	√	√	0.11	0.00	0.00	0.00	0.03	0.07	0.00	7.54	8.32	2.15	0.92
AT3G20360		√	0.15	0.13	0.36	0.20	0.28	1.23	0.65	1.89	2.55	1.23	1.64
AT3G21260	√		2.43	2.33	5.31	4.74	4.85	5.67	4.71	5.77	5.81	6.65	9.41
AT3G26320	√	√	3.06	2.90	3.27	2.90	2.38	7.29	3.44	9.35	8.69	11.87	7.97
AT3G27250	√		0.75	0.73	1.29	2.15	3.57	3.56	1.52	3.89	5.66	6.05	8.39
AT3G28910	√	√	11.34	12.90	13.93	18.13	18.69	22.96	18.81	26.27	24.66	34.20	30.29
AT3G44790		√	0.11	0.10	0.11	0.16	0.09	0.03	0.11	0.32	0.23	0.52	0.33
AT3G49845	√	√	0.00	0.00	0.00	0.00	0.09	0.16	0.07	0.44	1.87	1.55	0.46
AT3G50940	√	√	0.22	0.23	0.29	0.00	0.41	0.16	0.00	0.64	0.64	0.60	0.49
AT3G55890	√		0.34	1.07	0.57	0.43	0.53	2.56	0.22	3.45	4.57	2.67	5.93
AT3G55910	√	√	0.49	0.23	0.00	0.27	0.41	0.84	0.36	1.28	2.06	1.95	2.72
AT4G00680	√		0.08	0.13	0.00	0.55	0.28	0.10	0.33	0.88	1.99	2.31	0.56
AT4G22960	√	√	0.00	0.10	0.32	0.12	0.06	0.13	0.00	0.40	0.71	0.64	0.79
AT5G01600		√	0.00	0.00	0.00	0.00	0.00	0.23	0.00	0.28	0.38	0.32	0.23
AT5G03780	√	√	0.86	1.03	1.76	1.29	1.50	1.26	0.47	1.84	2.40	3.39	4.20
AT5G03860	√	√	0.00	0.03	0.11	0.47	0.06	0.55	0.91	1.16	2.85	1.27	1.74
AT5G07510		√	0.19	0.53	0.18	0.27	0.25	2.75	0.00	6.10	9.71	3.27	7.83
AT5G07530	√	√	0.00	0.00	0.07	0.47	0.09	0.19	0.00	4.21	2.92	0.60	1.48
AT5G13950		√	0.22	0.27	0.72	0.43	0.44	0.42	0.15	0.56	0.68	0.52	0.59
AT5G15800	√		1.08	0.83	0.65	0.82	1.25	2.01	0.54	2.53	2.81	6.01	6.33
AT5G16360	√	√	1.90	3.47	1.87	3.84	2.94	6.28	2.28	9.99	10.49	7.88	14.23
AT5G48720		√	0.19	0.13	0.83	0.16	0.13	0.32	0.15	0.44	0.45	0.52	0.46
AT5G60250	√	√	0.26	0.47	0.97	0.08	0.16	0.84	0.25	2.05	4.09	4.22	6.16

Down-regulated	ORF-based list	TAG-based list	Control			NIN				NAN			
			1406	9-5	5-15	9-5-9	9-5-13	9-5-14	9-5-16	9-5-10	9-5-11	9-5-12	9-5-15
AT1G75950		√	0.45	0.63	0.32	0.67	0.34	0.75	0.36	0.00	0.00	0.00	0.00
AT2G04852		√	0.22	0.20	0.36	0.35	0.28	0.19	0.11	0.00	0.08	0.04	0.07
AT3G05230		√	0.41	0.23	0.18	0.27	0.38	0.19	0.18	0.04	0.08	0.08	0.00
AT4G00480		√	0.41	0.57	0.18	0.27	0.28	0.23	0.33	0.20	0.19	0.20	0.16
AT4G04450	√		0.37	0.50	0.29	0.31	0.72	0.26	0.58	0.08	0.15	0.04	0.13
AT4G11230	√		0.22	0.23	0.25	0.20	0.31	0.32	0.25	0.08	0.00	0.00	0.07
AT4G14370		√	1.01	1.03	1.15	0.67	0.81	0.91	0.91	0.32	0.26	0.40	0.43
AT4G14400		√	0.52	0.97	0.57	0.35	0.44	0.23	0.22	0.00	0.00	0.00	0.20
AT4G18360		√	0.45	0.23	0.32	0.12	0.16	0.23	0.25	0.08	0.00	0.00	0.00
AT4G22517	√	√	11.79	10.10	6.71	8.93	13.90	8.13	4.57	3.21	2.06	3.15	4.49
AT4G39480		√	0.45	0.33	0.50	0.20	0.38	0.29	1.05	0.04	0.11	0.16	0.13
AT5G11670		√	0.86	1.17	3.12	0.51	1.16	0.75	0.58	0.00	0.00	0.00	0.00
AT5G16660	√	√	0.86	0.80	0.86	0.90	0.47	0.39	0.58	0.04	0.34	0.04	0.33
AT5G18360		√	1.23	0.93	0.18	0.43	0.69	0.49	0.65	0.32	0.41	0.36	0.26

NAN, nearest active analogues; NIN, nearest inactive analogues; TPM, transcripts per million sequences.

For seedlings containing ectopic expression constructs for the ORF corresponding to At5g15800, encoding AGAMOUS-like 2 also known as SEPALLATA 1, we noticed extremely early flowering, very similar to that observed due to over-expression of the related gene *Sepallata3* (Honma and Goto, 2001). This phenotype was not correlated with a change in iHR frequency. None of the other constructs used led to a clear and consistent phenotypic change.

Table 4 Recombination events scored in primary transformants with the different *CaMV35s* or *RPS5a* over-expresser lines of the indicated gene at 10 and 20 DPG

	CaMV35s (10 DPG)			*CaMV35s* (20 DPG)			*RPS5a* (10 DPG)			*RPS5a* (20 DPG)		
Gene	Spots	Plants with spots	Plants	Spots	Plants with spots	Plants	Spots	Plants with spots	Plants	Spots	Plants with spots	Plants
At1g11070	0	0	66	3	3	80	0	0	31	5	5	43
At1g55380	4	4	115	15	14	116	1	1	85	26	23	111
At1g65300	1	1	114	4	4	139	n.d.	n.d.	n.d.	0	0	6
At1g66300	1	1	86	33	22	109	4	4	58	43	35	118
At2g18193	3	2	26	4	4	53	2	2	64	11	11	75
At2g22630	0	0	31	16	13	61	1	1	45	12	10	60
At2g27880	0	0	39	8	6	64	0	0	63	4	4	78
At3g20360	0	0	15	3	3	39	0	0	30	11	9	64
At3g21260	0	0	43	2	2	71	1	1	53	18	16	89
At3g26320	0	0	26	2	1	40	0	0	26	1	1	50
At3g27250	0	0	11	3	2	32	0	0	7	2	2	25
At3g28910	0	0	15	3	3	28	0	0	9	6	6	27
At3g49845	n.d.	n.d.	n.d.	0	0	11	3	3	84	4	4	81
At3g50940	0	0	13	0	0	63	1	1	41	2	2	75
At3g55890	0	0	47	1	1	70	0	0	33	0	0	63
At3g55910	1	1	80	3	2	132	0	0	4	0	0	13
At4g00680	1	1	70	7	7	97	1	1	51	5	5	54
At4g22960	1	1	83	6	6	100	1	1	45	1	1	75
At5g03780	0	0	74	1	1	88	n.d.	n.d.	n.d.	0	0	3
At5g03860	0	0	42	6	5	57	1	1	38	4	4	70
At5g07530	0	0	49	7	7	68	1	1	61	9	7	79
At5g15800	4	4	83	1	1	110	3	3	90	8	4	108
At5g16360	0	0	68	5	4	74	0	0	36	5	4	83
At5g48720	1	1	53	1	1	85	0	0	87	14	13	106
At5g60250	0	0	8	1	1	43	2	2	40	1	1	82
Empty vector	0	0	40	0	0	64	0	0	10	0	0	25
1406	2	2	270	21	19	415	2	2	270	21	19	415

DPG, days after germination.

Discussion

In this work, we wanted to investigate the molecular mechanism underlying the 200- to 300-fold increase in iHR frequency found to be evoked in *Arabidopsis* plants by the expression of the artificial transcription factor VP16-HRU. This factor was discovered in previous work, employing a powerful genome interrogation technique using 3F ZF-ATFs in order to find mutant plants with increased iHR (Lindhout *et al.*, 2006). With only a single mutant found, the task to clearly identify possible causal changes in gene expression due to VP16-HRU expression proved to be too difficult; experiments with microarrays showed a multitude of differentially expressed genes (Lindhout, 2008), as in fact can be expected when expressing a 3ZF-ATF with potentially about 1000 cognate 9-bp DNA-binding sites within the *Arabidopsis* genome. In the present work, it is demonstrated that generating the nearest inactive and active neighbours of a particular ZF-ATF, a so-called NIN/NAN approach where all ZF-ATF factors share a particular 2F domain, can in combination with a deep sequencing strategy provide meaningful insights into the causal changes related to a ZF-ATF-induced phenotype. In the case of iHR induction, we obtained evidence supporting a model where several plant gene products rather than a single master switch lead to increased iHR.

The idea behind the NIN/NAN strategy was inspired by the fact that ZF–DNA recognition is not completely specific (Carroll *et al.*, 2006; Segal *et al.*, 1999). While this can present a problem for the design of truly site-specific PZF domains, the lack of true specificity can be a benefit in relation to genome interrogation with 3F ZF-ATFs; it should be possible that different but highly related PZF domains can evoke a similar phenotype of interest by sharing interactions with one or more particular 9-bp DNA sequences. In cases where practical reasons had set limits to a ZF-ATF-containing mutant population, a generic strategy to generate and test a dedicated series of different structural analogues of an interesting ZF-ATF would alleviate the problem of getting to know possible causal changes of interest; although generated afterwards, different NIN/NAN analogues should then pave the way for discriminatory screening of transcriptome data.

Generating a series of 40 3F-ZF-ATFs with the first two ZFs identical to those in the original ZF-ATF (VP16-HRU) that triggered an extremely high frequency of iHR and expressing those in plants clearly demonstrated that the idea of NIN/NAN analogues was valid. In addition to VP16-HRU, the original 3F ZF-ATF with the 3F combination of GNN-recognizing ZFs 9-5-15, we found that also the full GNN-type PZF domains 9-5-10, 9-5-11 and 9-5-12 were inducing iHR to a similarly high level when combined with the VP16 domain (Figure 1). Besides the high iHR frequency,

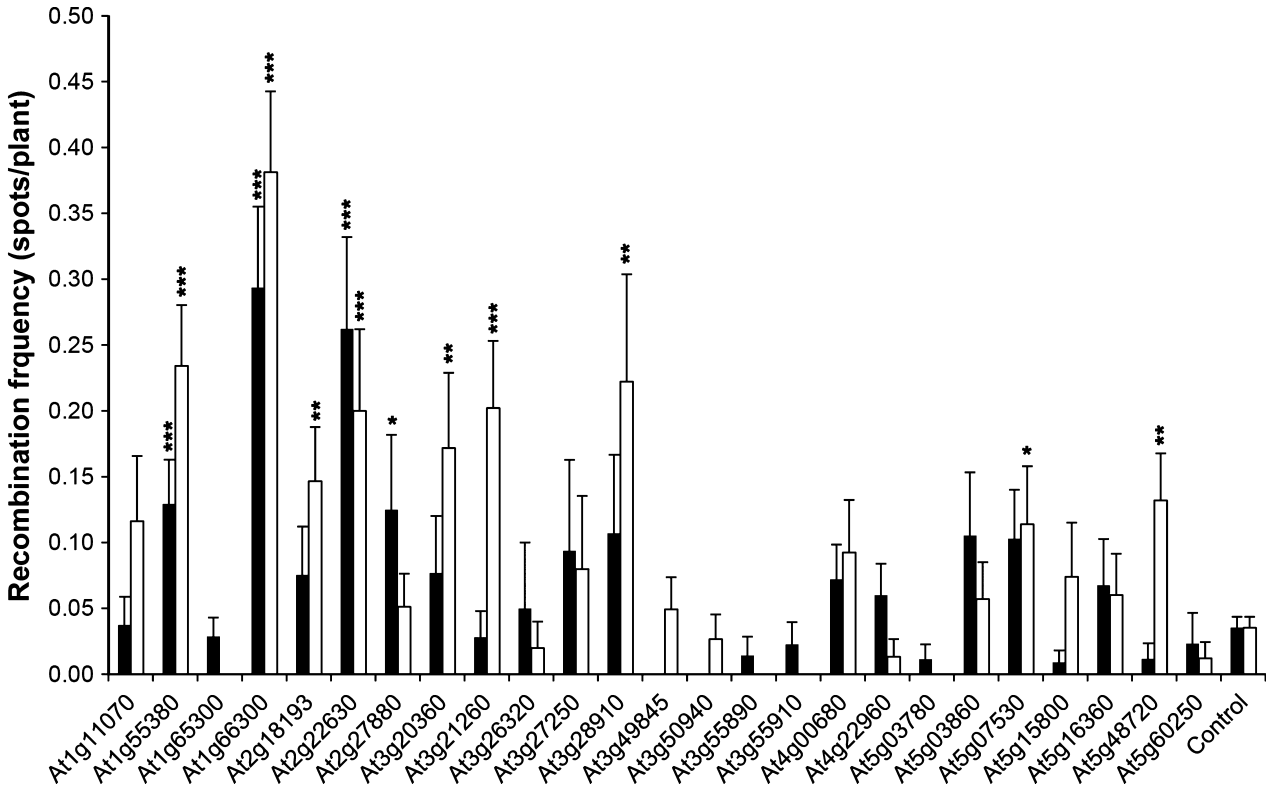

Figure 3 Homologous recombination frequency, quantified as the number of blue spots per seedling at 20 DPG with the selected genes driven by both promoters (*CaMV35S* in black bars and *RPS5A* in white bars). Error bars represent SEM. One, two and three asterisks represent *P*-value <0.05, <0.01 and <0.001 compared with control, respectively.

increased at least 100-fold compared with figures seen for inactive constructs, further phenotypic traits were normal. With in total four NANs with full GNN-type ZF domains, the most similar NINs were also picked within this group. Altogether, testing 40 different 9-5-X constructs clearly proved specificity of iHR induction, with NANs only found when 'X' was a GNN-type ZF (Figure 1). For future NIN/NAN designs of the same kind, extension of the founding 2F domain with a third ZF of the same type (GNN, ANN, CNN or TNN) as the third finger present in the original factor should be sufficient to find suitable NIN/NANs.

The number of transcripts that can be regarded as differentially regulated depends on the particular criteria set for an experiment. In our case, on average about 1200 genes can be regarded as differentially regulated for each particular NAN transcriptome (Figure 2). About half of them were up-regulated, but only 68 of them were shared between all NAN transcriptomes. Having access to four NAN transcriptomes thus reduced the complexity enormously, but further experimentation with 68 genes was still too demanding. By including a requirement that finally up to all four NIN transcriptomes do not exhibit a larger change in transcript level for a particular gene than found in any of the NAN transcriptomes, the number of genes was gradually more than halved. Although this decrease in gene number due to employing NIN data is relatively modest, it served very well to get towards a more manageable series for further experimentation. Of course, as mentioned in the results section above, employing NIN data has a concomitant risk of losing the full picture upon a small series of plant genes that trigger a phenotype of interest, as it can happen that one or more of these genes are differentially expressed in one or more NIN lines. However, an endogenous master switch should still be found. Whatever the mechanism underlying enhanced iHR in our original VP16-HRU-expressing plants, we considered it most important that the genes to be investigated by means of ectopic expression were selected via a procedure that was free from criteria that might cause experimental bias due to reasons beyond the given facts of the data set.

After testing the effects of ectopic expression of almost all ORFs corresponding to the final selection of genes via the *RPS5A* and *CaMV35S* promoters, it has to be concluded that an endogenous master switch leading to enhanced iHR at a level comparable to NAN expression has not been identified. This does of course not fully exclude the possibility that an endogenous master switch exists; it can for instance be envisaged that a relatively minor change in the expression of one of the components of a signal transduction pathway can finally lead to an amplified response, in this case a hugely enhanced iHR frequency. In our case, such a master switch could then be identified when all genes with less outspoken but still consistent changes in NAN-expressing plant lines will be experimentally addressed for their potential to induce iHR events, thus also including genes exhibiting less than twofold changes in expression levels. Obviously, such an approach will stretch far beyond practical limitations.

Irrespective of the question whether a natural iHR master switch exists or not, the picture that emerged after our experiments is fully compatible with a model where NAN-type ZF-ATFs are potent exotic master switches for iHR; several individual genes that are directly or indirectly under NAN control are able to contribute to a much higher iHR frequency. It is

therefore tempting to speculate that concerted action of these genes might trigger the extremely high iHR frequency found in NAN-expressing plants. However, getting definitive experimental proof of this idea is quite demanding as it is likely to require mimicking the expression characteristics of the different genes as they occur in plant lines expressing NAN-type ZF-ATFs. Moreover, it remains possible that some of those genes not demonstrated to be linked with iHR may still be required to reach the full iHR potential.

Whatever the precise mechanism behind the very high iHR frequencies found in NAN-expressing plant lines, it is evident that genome interrogation followed by a NIN/NAN strategy resulted in the identification of ORFs that when ectopically expressed can lead to up to 10-fold higher iHR frequency. Compared to published data regarding the effects of individual genes or mutations upon iHR frequencies (Bagherieh-Najjar *et al.*, 2005; Endo *et al.*, 2006; Gherbi *et al.*, 2001; Heitzeberg *et al.*, 2004; Knoll *et al.*, 2012; Molinier *et al.*, 2005; Takeda *et al.*, 2004; Yin *et al.*, 2009), a 10-fold increase by a single ORF is among the highest changes reported. The ORFs most clearly linked with enhanced iHR correspond to At1g55380, At1g66300 and At2g22630. All three ORFs led to highly significant changes in iHR frequency ($P < 0.001$) under control of the *RPS5A* as well as the *CaMV35S* promoter. Remarkably, none of the proposed functions for the encoded proteins (At1g55380, a cysteine/histidine-rich C1 family protein; At1g66300, an F-box/RNI-like/FBD-like domains-containing protein; At2g22630, transcription factor Agamous-like 17) provide a direct like with a role in DNA metabolism. In fact, this also holds for the other ORFs that were selected (Table 3). Only At5g48720 (X-ray-induced transcript) was reported to interact with the HR machinery and to display transcriptional induction in response to DNA double-strand breaks in the ATM-dependent pathway (Dean *et al.*, 2009). Yet, in our experiments, iHR induction with At5g48720 was modest (*RPS5A* promoter, $P < 0.01$) or absent (CaMV35S promoter).

Interestingly, transcriptome analyses of gamma-ray-irradiated plants (Cools *et al.*, 2011; Culligan *et al.*, 2006; Ricaud *et al.*, 2007; Yoshiyama *et al.*, 2009) were shown to exhibit among others enhanced levels of mRNAs corresponding to At4g22960, At5g03780, At5g48720 and At5g60250, all found within our selected series. However, apart from At5g48720 just mentioned above, an experimental link with iHR frequencies was not detected in our experiments. Considering other studies dealing with enhanced iHR due to different types of mutations (Bagherieh-Najjar *et al.*, 2005; Endo *et al.*, 2006; Gherbi *et al.*, 2001; Heitzeberg *et al.*, 2004; Knoll *et al.*, 2012; Molinier *et al.*, 2005; Takeda *et al.*, 2004; Yin *et al.*, 2009), overlap with our selection appears to be lacking. A plethora of different genes has thus been linked with changed iHR frequencies, but the VP16-HRU pathway seems to be rather unique with its very high iHR frequency likely to be caused by a concerted action of several iHR-inducing genes.

One of the remarkable aspects of the iHR phenotypes observed is that the actual iHR events are found in more matured cells, while the expression of NANs and other ZF-ATFs via the *RPS5A* promoter is expected to occur predominantly in early embryonic phase and in meristems (Lindhout *et al.*, 2007; Weijers *et al.*, 2001). Hence, when using *RPS5A* promoter-mediated expression, almost all iHR events do in fact occur when NAN expression levels drop. Experiments where VP16-HRU expression was driven by the constitutively active *CaMV35S* promoter did not show further induction of iHR frequency (Lindhout and Van der Zaal, unpublished data). Altogether, a picture emerges as if crucial epigenetic marks or gene expression patterns can be established in immature cells via the expression of ZF-ATFs, in the present case still facilitating iHR events when cells mature. The fact that expression of individual genes could also lead to enhanced iHR can be regarded as an important proof that iHR-inducing NANs act by establishing a gene expression pattern rather than marks on the GU-US reporter locus itself.

In conclusion, even when practical limitations to raising large numbers of ZF-ATF-expressing individuals exist, it can be stated that a dedicated strategy aimed at generating analogues of active ZF-ATFs can provide a breakthrough for mutant analysis. Genome interrogation can therefore be regarded as a valuable tool, not only for finding novel mutants, but also for providing insights in the molecular genetic events causal for a phenotype of interest.

Experimental procedures

Constructs and molecular cloning

The plant expression vector pGPTV-KAN (Becker *et al.*, 1992) was modified to obtain the vector pRN, with a unique *Not*I site in between the *RPS5A* promoter and a *t-nos* terminator (Lindhout, 2008). For *CaMV35S* promoter-mediated expression, the RPS5A promoter was replaced by an 874-bp *Xba*I-*Not*I fragment containing the constitutive *CaMV35S* promoter from pINDEX3 (positions 13-879) (Ouwerkerk *et al.*, 2001) by PCR amplification and restriction enzyme digestion, thus generating plasmid p35SN.

Phusion™ DNA high-fidelity polymerase (Finnzymes, Espoo, Finland) was used to obtain PCR fragments representing ORFs of selected genes, mostly using lambda ZAP cDNA libraries prepared from mRNA derived from *Arabidopsis* roots, some from single-stranded cDNA derived from mRNA isolated from seedlings expressing VP16-HRU. The primers are listed in Table S2. PCR products were sequence-verified (Macrogen, Seoul, Korea) and cloned via their PCR-introduced restriction sites into a derivative of pSKN-*Sgr*AI (Neuteboom *et al.*, 2006), harbouring the matching restriction sites in between *Not*I sites. Some cDNA sequences obtained had one or more SNPs compared with the Col-0 TAIR10 annotation (data not shown), but these rarely caused a change in encoded amino acids and never any frame shifts or stops. Most ORFs acquired an AMV translational enhancer as mentioned in text. Subsequently, sequences were transferred as *Not*I fragments to vectors pRN and p35SN, the proper orientation was determined by restriction enzyme digestion, and plasmids were introduced into *Agrobacterium tumefaciens* strain AGL1 (Lazo *et al.*, 1991) via triparental mating (Ditta *et al.*, 1980). For unknown reasons, PCR products representing At1g11070 and At2g27880 were rather recalcitrant to cloning. In these cases, flanking *Not*I sites were introduced via additional primers resulting in loss of the AMV leader sequence.

Plant transformation and homologous recombination assay

Arabidopsis plants of iHR reporter line 1406 (Gherbi *et al.*, 2001) were transformed with the various expression constructs via floral dip (Clough and Bent, 1998). The primary transformants (T1 seedlings) were selected on MA medium (Masson and Paszkowski, 1992) lacking sucrose and containing 50 mg/L kanamycin, 100 mg/L timentin and 100 mg/L nystatin. At 10 and 20 DPG, primary transformants were histochemically stained for GUS activity as described previously (Lindhout *et al.*, 2006). The number of spots per seedling was counted to determine the number of recombination events. Statistical analysis was per-

formed using an ANOVA test against the sum of all empty vector controls and the 1406 line within the SPSS package (SPSS Inc., Chicago, IL).

Experimental set-up for NGS

Samples of seeds were surface-sterilized with 70% ethanol and vernalized on moisturized Whatmann 3MM paper for 4 days at 4 °C after which they were placed overnight at room temperature. In total, for each genotype, 200 germinated seeds were transferred to 20 small pots (ten seeds per pot) containing a sand–soil mixture in a fully randomized set-up. Some additional plants were potted to confirm the enhanced HR rate of the iHR phenotype of each line by GUS staining. Prior to harvesting, each individual pot was inspected for plants that were lacking behind in development or appeared less healthy. In such a relatively rare case, the entire pot containing ten plants was removed from the experiment. From the remaining pots, aerial parts of seedlings were harvested at 20 DPG in a fully randomized fashion. Per construct represented, the material was pooled and RNA was extracted using the Qiagen (Hilden, Germany) RNAeasy plant kit with on column DNAse treatment. The initial RNA quality and purity was investigated using the NanoDrop resulting in an average RNA yield of 257.9 $\pm$ 21.9 ng/μL, 260/280 ratio of 2.18 $\pm$ 0.01 and an 260/230 ratio of 2.27 $\pm$ 0.12.

Illumina DGE

Before initiating the library preparation, the quality of all RNA samples was checked using an Agilent plant RNA Nano Lab on Chip assay resulting in an average RNA integrity (RIN) value of 8.90 $\pm$ 0.10. The library preparation was performed using 1.5 μg of total RNA per sample using the standard Illumina DGE NlaIII profiling protocol. Each created library was loaded on a full lane of a GAIIx flow cell and sequenced by the Leiden Genome Technology Center (LGTC).

The obtained total number of RAW reads per sample (average 28.4 $\pm$ 2.5 million) was processed using the CLC genomics workbench into expression values for each tag or corresponding gene in TPM using the following settings: include reverse, downstream tag, NlaIII restriction site, all restriction types, allow single substitutions/insertions/deletions, annotate all, prefer high-priority mutant and with the priority of tags as 3′, 5′, Int, 3′(-), 5′(-), Int(-). The total number of tags per sample that could be assigned to a location within the predicted *Arabidopsis* transcriptome (TAIR 10 release) was on average 28.1 $\pm$ 2.4 million.

Acknowledgements

This work was funded by the Dutch Technological Top Institute Green Genetics (TTI GG, grant number 3CC041RP) and by Rijk Zwaan Breeding Company (The Netherlands). We gratefully acknowledge discussions with Drs Rob Dirks, Kees van Dun and Cilia Lelivelt (Rijk Zwaan) and the technical support of Gary Strunk (BSc), Ward de Winter and Jan Vink.

References

Bagherieh-Najjar, M.B., de Vries, O.M., Hille, J. and Dijkwel, P.P. (2005) Arabidopsis RecQl4A suppresses homologous recombination and modulates DNA damage responses. *Plant J.* **43**, 789–798.

Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.* **20**, 1195–1201.

Beltran, A., Liu, Y., Parikh, S., Temple, B. and Blancafort, P. (2006) Interrogating genomes with combinatorial artificial transcription factor libraries: asking zinc finger questions. *Assay Drug Dev. Technol.* **4**, 317–347.

Bleuyard, J.Y., Gallego, M.E. and White, C.I. (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair*, **5**, 1–12.

Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and *in vitro* testing of zinc finger nucleases. *Nat. Protoc.* **1**, 1329–1341.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Cools, T., Iantcheva, A., Weimer, A.K., Boens, S., Takahashi, N., Maes, S., Van den Daele, H., Van Isterdael, G., Schnittger, A. and De Veylder, L. (2011) The *Arabidopsis thaliana* checkpoint kinase WEE1 protects against premature vascular differentiation during replication stress. *Plant Cell*, **23**, 1435–1448.

Culligan, K.M., Robertson, C.E., Foreman, J., Doerner, P. and Britt, A.B. (2006) ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J.* **48**, 947–961.

Datla, R.S., Bekkaoui, F., Hammerlindl, J.K., Pilate, G., Dunstan, D.L. and Crosby, W.L. (1993) Improved high-level constitutive foreign gene expression in plants using an AMV RNA4 untranslated leader sequence. *Plant Sci.* **94**, 139–149.

Dean, P.J., Siwiec, T., Waterworth, W.M., Schlögelhofer, P., Armstrong, S.J. and West, C.E. (2009) A novel ATM-dependent X-ray-inducible gene is essential for both plant meiosis and gametogenesis. *Plant J.* **58**, 791–802.

Ditta, G., Stanfield, S., Corbin, D. and Helinski, D.R. (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. *Proc. Natl Acad. Sci. U.S.A.* **77**, 7347–7351.

Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F. 3rd. (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **276**, 29466–29478.

Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B. and Barbas, C.F. 3rd. (2005) Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **280**, 35588–35597.

Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K., Abe, K., Ichikawa, H., Valentine, L., Hohn, B. and Toki, S. (2006) Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. *EMBO J.* **25**, 5579–5590.

Gherbi, H., Gallego, M.E., Jalut, N., Lucht, J.M., Hohn, B. and White, C.I. (2001) Homologous recombination in planta is stimulated in the absence of Rad50. *EMBO Rep.* **2**, 287–291.

Heitzeberg, F., Chen, I.P., Hartung, F., Orel, N., Angelis, K.J. and Puchta, H. (2004) The Rad17 homologue of Arabidopsis is involved in the regulation of DNA damage repair and homologous recombination. *Plant J.* **38**, 954–968.

Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525–529.

Knoll, A., Higgins, J.D., Seeliger, K., Reha, S.J., Dangel, N.J., Bauknecht, M., Schröpfer, S., Franklin, F.C. and Puchta, H. (2012) The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis. *Plant Cell*, **24**, 1448–1464.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation–competent arabidopsis genomic library in agrobacterium. *Biotechnology (N Y)*, **9**, 963–967.

Lee, H.J., Kim, E. and Kim, J.S. (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* **20**, 81–89.

Lee, J., Hirsh, A.S., Wittner, B.S., Maeder, M.L., Singavarapu, R., Lang, M., Janarthanan, S., McDermott, U., Yajnik, V., Ramaswamy, S., Joung, J.K. and Sgroi, D.C. (2011) Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS ONE* **6**, e21112.

Lieberman-Lazarovich, M. and Levy, A.A. (2011) Homologous recombination in plants: an antireview. *Methods Mol. Biol.* **701**, 51–65.

Lindhout, B.I. (2008) *Regulatory DNA binding peptides as novel tools for plant functional genomics*. PhD. Thesis, Leiden University.

Lindhout, B.I., Pinas, J.E., Hooykaas, P.J. and van der Zaal, B.J. (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J.* **48**, 475–483.

Lindhout, B.I., Fransz, P., Tessadori, F., Meckel, T., Hooykaas, P.J. and van der Zaal, B.J. (2007) Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res.* **35**, e107.

Markmann-Mulisch, U., Wendeler, E., Zobell, O., Schween, G., Steinbiss, H.H. and Reiss, B. (2007) Differential requirements for RAD51 in *Physcomitrella patens* and *Arabidopsis thaliana* development and DNA damage repair. *Plant Cell*, **19**, 3080–3089.

Masson, J. and Paszkowski, J. (1992) The culture response of *Arabidopsis thaliana* protoplasts is determined by the growth conditions of donor plants. *Plant J.* **2**, 829–833.

Meyers, B.C., Lee, D.K., Vu, T.H., Tej, S.S., Edberg, S.B., Matvienko, M. and Tindell, L.D. (2004) *Arabidopsis* MPSS. An online resource for quantitative expression analysis. *Plant Physiol.* **135**, 801–813.

Molinier, J., Oakeley, E.J., Niederhauser, O., Kovalchuk, I. and Hohn, B. (2005) Dynamic response of plant genome to ultraviolet radiation and other genotoxic stresses. *Mutat. Res.* **571**, 235–247.

Neuteboom, L.W., Lindhout, B.I., Saman, I.L., Hooykaas, P.J. and van der Zaal, B.J. (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem. Biophys. Res. Commun.* **339**, 263–270.

Ouwerkerk, P.B., de Kam, R.J., Hoge, J.H. and Meijer, A.H. (2001) Glucocorticoid-inducible gene expression in rice. *Planta*, **213**, 370–378.

Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, **252**, 809–817.

Ricaud, L., Proux, C., Renou, J.P., Pichon, O., Fochesato, S., Ortet, P. and Montané, M.H. (2007) ATM-mediated transcriptional and developmental responses to gamma-rays in Arabidopsis. *PLoS ONE*, **2**, e430.

Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F. 3rd. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl Acad. Sci. U.S.A.* **96**, 2758–2763.

Sera, T. (2009) Zinc-finger-based artificial transcription factors and their applications. *Adv. Drug Deliv. Rev.* **61**, 513–526.

Shaked, H., Melamed-Bessudo, C. and Levy, A.A. (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc. Natl Acad. Sci. U.S.A.* **102**, 12265–12269.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., Choi, V.M., Rock, J.M., Wu, Y.Y., Katibah, G.E., Zhifang, G., McCaskill, D., Simpson, M.A., Blakeslee, B., Greenwalt, S.A., Butler, H.J., Hinkley, S.J., Zhang, L., Rebar, E.J., Gregory, P.D. and Urnov, F.D. (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature*, **459**, 437–441.

Takeda, S., Tadele, Z., Hofmann, I., Probst, A.V., Angelis, K.J., Kaya, H., Araki, T., Mengiste, T., Mittelsten Scheid, O., Shibahara, K., Scheel, D. and Paszkowski, J. (2004) BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. *Genes Dev.* **18**, 782–793.

Weijers, D., Franke-van Dijk, M., Vencken, R.J., Quint, A., Hooykaas, P. and Offringa, R. (2001) An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development*, **128**, 4289–4299.

Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., Fu, F., Porteus, M.H., Dobbs, D., Voytas, D.F. and Joung, J.K. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.* **1**, 1637–1652.

Yin, H., Zhang, X., Liu, J., Wang, Y., He, J., Yang, T., Hong, X., Yang, Q. and Gong, Z. (2009) Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in Arabidopsis. *Plant Cell*, **21**, 386–402. Epub 2009 Feb 24.

Yoshiyama, K., Conklin, P.A., Huefner, N.D. and Britt, A.B. (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc. Natl Acad. Sci. U.S.A.* **106**, 12843–12848.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Homologous recombination frequency, quantified as the number of blue spots per seedling at 10 DPG with the selected genes driven by both promoters (*CaMV35S* in black bars, *RPS5A* in white bars).

Table S1 Description of selected candidate genes.

Table S2 Primers for the amplification of the target genes and the gene sizes.

