



## OPEN

## Optimized tuning of TALEN specificity using non-conventional RVDs

## SUBJECT AREAS:

MOLECULAR  
ENGINEERING  
GENETIC ENGINEERING

Alexandre Juillerat, Coline Pessereau, Gwendoline Dubois, Valérie Guyot, Alan Maréchal, Julien Valton, Fayza Daboussi, Laurent Poirot, Aymeric Duclert &amp; Philippe Duchateau

COLLECTIS S.A., 8 Rue de la Croix Jarry, 75013 Paris, France.

Received  
20 November 2014Accepted  
22 December 2014Published  
30 January 2015Correspondence and  
requests for materials  
should be addressed to  
A.J. (alexandre.  
juillerat@collectis.com)  
or P.D. (philippe.  
duchateau@collectis.  
com)

A key feature when designing DNA targeting tools and especially nucleases is specificity. The ability to control and tune this important parameter represents an invaluable advance to the development of such molecular scissors. Here, we identified and characterized new non-conventional RVDs (ncRVDs) that possess novel intrinsic targeting specificity features. We further report a strategy to control TALEN targeting based on the exclusion capacities of ncRVDs (discrimination between different nucleotides). By implementing such ncRVDs, we demonstrated in living cells the possibility to efficiently promote TALEN-mediated processing of a target in the HBB locus and alleviate undesired off-site cleavage. We anticipate that this method can greatly benefit to designer nucleases, especially for therapeutic applications and synthetic biology.

The past few years have seen the emergence of novel and extremely robust DNA targeting platforms namely the transcription activator like effectors (TALE) and the clustered regularly interspaced short palindromic repeats (CRISPR)<sup>1–3</sup>. The possibility to control and finely tune the targeting specificity of such platforms represents a key issue. Regarding TALEs, DNA binding core of this class of proteins is composed of 33 to 35 amino acid-long repeats arranged in a tandem array. Each of the individual repeated modules differs at two essential positions (12 and 13), the so-called repeat variable diresidue (RVD) and contacts one single nucleotide on one strand of the DNA target in a N-terminus to C-terminus direction (in line with 5' to 3' of the target)<sup>4,5</sup>. Structural, biochemical and functional studies underlined that the nature of the amino acid located at position 13 was responsible for the specificity of recognition by contacting in the major groove the top DNA strand base<sup>6–8</sup>. The amino acid in position 12 participates, on the other side, to the stabilization of the repeated units.

The current TALE scaffold turned out to be a very robust DNA targeting platform for molecular tools. Today, only 4 DNA targeting modules (RVDs) are mainly implemented and used by most researchers (NI, HD, NN, and NG). Recently a few studies have started to highlight the benefit to implement other natural RVDs<sup>9,10</sup>. However, the naturally found RVDs only explore a limited fraction (~5%) of the possible diversity repertoire at these two key positions.

As naturally found RVDs present various patterns of specificity on the four nucleotides, we hypothesized that exploring the 400 combinations of the two key amino acids (position 12 and 13) could allow pinpointing non-conventional RVDs (ncRVD) with novel alternative specificity patterns. To identify such ncRVDs, we generated and screened collections of TALEN containing alternative combinations of amino acids at position 12 and 13 at defined positions of an array. Among the thousands TALEN/targets combinations tested (~18,000), we identified ncRVDs that presented novel valuable specificity features (exclusion properties). The additional levels of specificity control offered by these alternative DNA targeting modules will clearly represent a major advantage to alleviate off-site targeting. Off-site targets are defined as sequences within the genome that contain a few mismatches relative to the targeted sequence of interest. When considering an intended target and a related off-site sequence, the adequate implementation of ncRVDs that discriminate between the nucleotides present at the position of the mismatches (exclusion properties) will prevent the TALEN activity at the off-site sequences while allowing processing of the desired locus. As proof of concept, we demonstrated the potential of ncRVDs incorporation in TALE arrays by generating TALEN able to discriminate, in mammalian cells, between two closely related loci in the HBB (in the vicinity of the mutation responsible for the sickle cell anaemia) and HBD (off-site sequence, 94% of identity with HBB) genes.

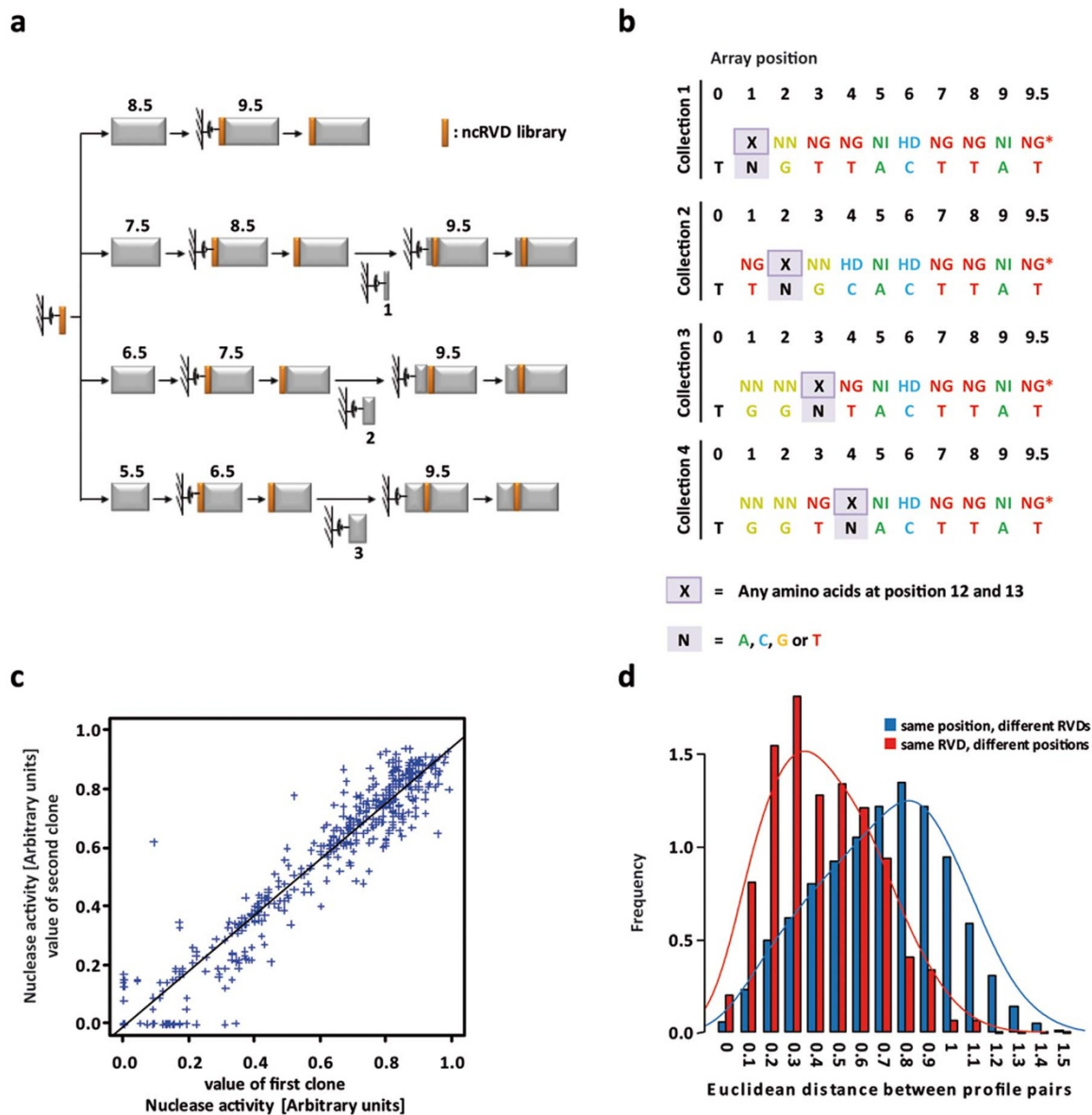


## Results

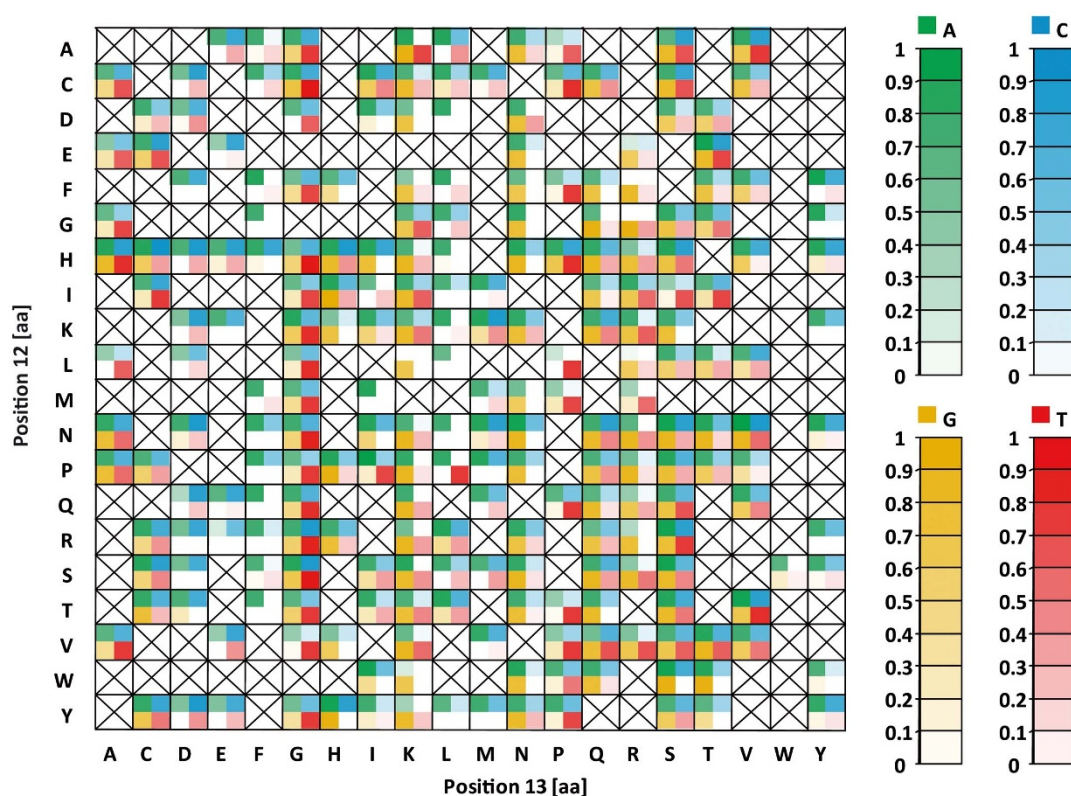
**Experimental setup and collection screening.** To systematically explore the landscape of RVD specificity and stringency we implemented a library-based high-throughput screen of TALE nuclease (TALEN) containing non-conventional RVDs (ncRVDs)<sup>11–13</sup>. First, a library of degenerated RVDs randomized at position 12 and 13 was created by short overlap PCR (SOE) using the conventional NNK codon degeneracy that encodes the complete set of natural amino acids. This RVD library was further incorporated at different position of TALE arrays using a solid phase assembly strategy (Fig. 1a)<sup>14</sup>. We

focused in particular on the first four positions of the array (Fig. 1b) as we<sup>12</sup> and others<sup>15</sup> demonstrated that this N-terminal section of the array tends to present a higher specificity than the C-terminal section.

The four TALEN collections were assayed against their respective targets containing the four nucleotides (A, C, G and T) at the adequate position, regarding the localization of the ncRVD in the TALE array (Fig. 1b). Altogether, about 18,000 TALEN/targets combinations were assayed (~4200 to 4600 per collection). For each collection, we selected approximately hundred clones (101 clones for position 1, 105 for position 2, 136 for position 3 and 128 for



**Figure 1 | Design of experimental setup and validation experiments.** (a). Graphical representation of the library assembly process. The randomized RVD is represented in orange and pre-assembled blocs in grey. The size of each bloc is indicated. The solid support is composed of streptavidin coated magnetic beads. (b). Schematic representation of the libraries of TALE and targets used in the yeast activity assay. This nuclease activity assay relies on the single strand annealing (SSA) pathway used after the generation, by the TALEN, of a double strand break (DSB) in the target sequence. Target sites were designed for homodimeric TALEN. X represents any combination of the 20 amino acids (encoded by the NNK codon) at position 12 and 13. N represents any of the four A, C, G and T base. Libraries are composed of arrays containing 9.5 repeats. (c) Comparison of the activity between clones encoding the same TALE array. For each TALEN for which two clones encoding the same protein were obtained, the activity was compared on all nucleotides. Each point corresponds to a pair of clones measured on a given nucleotide (first and second clone are arbitrarily fixed). (d) Comparison of the distribution of distances between profiles of TALEN pairs sharing the same RVD at different positions, and TALEN pairs having different RVDs at the same position. For each pair of TALEN, the Euclidean distance was computed between their profiles on the four nucleotides. Pairs for the blue distribution have different RVDs at the same position (from the same collection), while pairs for the red distribution have the same RVD at different positions (from different collections). The curves are density curves corresponding to the histograms.



**Figure 2 | Analysis of global nuclease activity in function of amino acid 12 and 13 identity.** Heatmap showing the nuclease activity of identified combinations of amino acids at position 12 and 13 of an RVD on targets containing the four possible nucleotides. The activity was averaged on all four collections. A cross indicates that the combination of amino acids was not found in the present analysis of TALE array libraries, and therefore, the activity is unknown. Activity is represented on a scale from 0 (no nuclease activity) to 1 (strong nuclease activity) by increasingly darker colors (arbitrary units).

position 4) that displayed a high nuclease activity in the yeast assay on at least one of the four targets (containing A, C, G and/or T). All these clones were individually re-assayed to ensure consistency of measurements. We further sequenced the validated clones and overall, found that all 20 amino acids could be recovered at either position 12 or 13. However, the frequency of apparition of each amino acid was highly variable, with predominance for naturally occurring amino acids (Supplementary Fig. 1a and 1b). We next monitored the variability of the nuclease activity for a given RVD (same position in the array), taking also into account the codon degeneracy. The mean activity and standard deviation were computed, revealing a low variability, altogether validating the experimental setup (Fig. 1c).

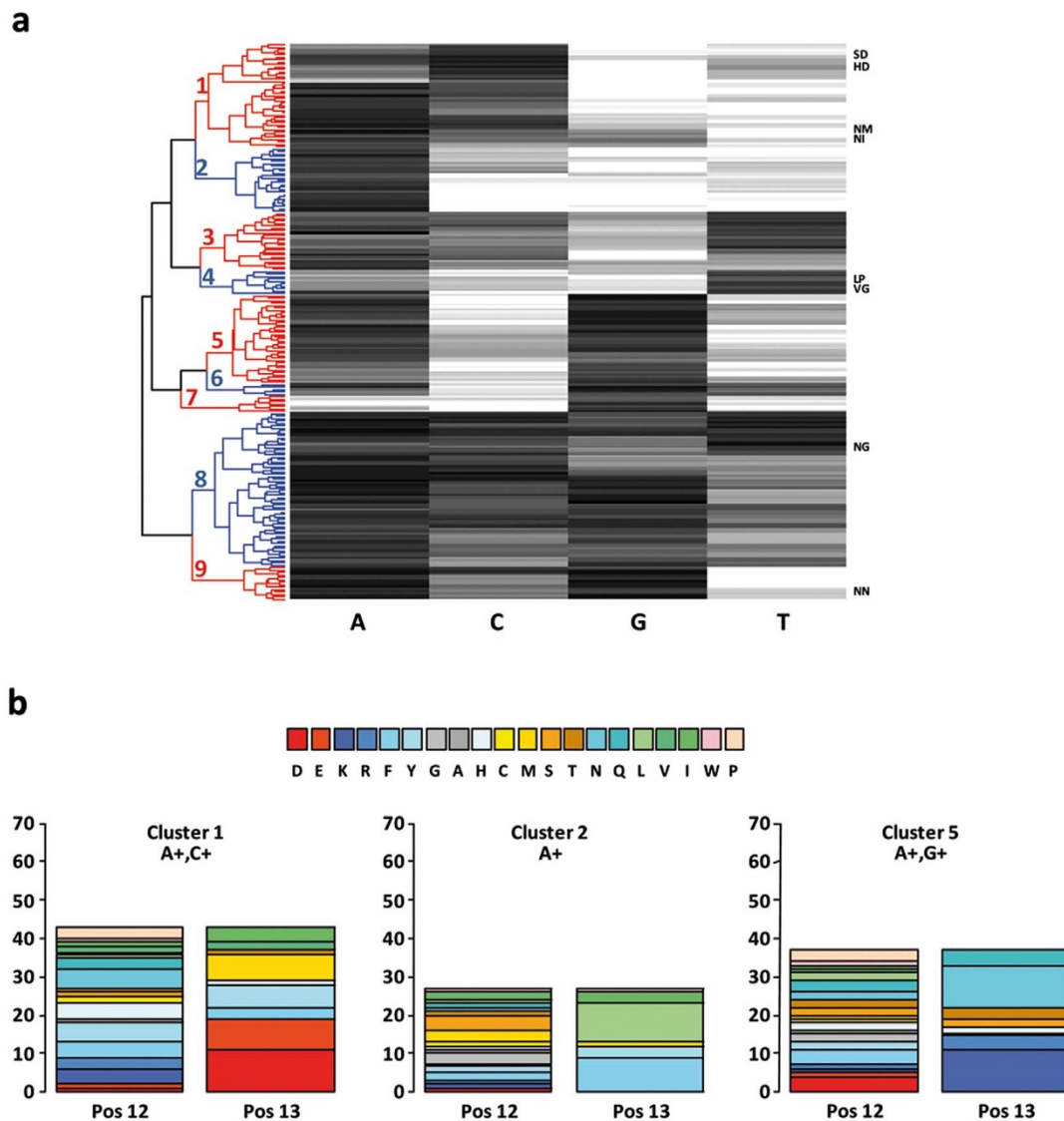
**Variability by position.** The experimental setup we used allowed studying several parameters including the effect of the amino acids identities but also position in the array. To investigate the effect of these two parameters, we computed the variability on the cleavage profiles due to (i) the position (1 to 4) in the array, keeping the amino acids in position 12 and 13 fixed and (ii) the identity of amino acids in position 12 and 13, keeping the position 1 to 4 fixed. We thus calculated the euclidean distance between all pairs of profiles for the two sets (position or identity) which gave a distribution of distances (Fig. 1d). The comparison of both distributions clearly pointed out that the nature of the amino acids in position 12 and 13 has a much more pronounced impact on the nuclease activity compared to its position in the array, as demonstrated by the shift to the left (smaller distances) for the position distribution (Fig. 1d). Noteworthy, this feature was also previously observed in our recent study focused exclusively on the activity and specificity the four mostly used RVDs (NI, HD, NN and NG)<sup>12</sup>.

**Analysis of the cleavage profiles of ncRVDs on the 4 nucleotides.** For each collection, we then calculated heatmaps of the nuclease

activity as function of the amino acid found in position 12 and 13 (Supplementary Fig. 2, 3, 4 and 5). As we previously showed that the identity of the amino acids in position 12 and 13 is of major importance compared to the position in the array, we additionally calculated a global heatmap, where nuclease activity values were averaged on all positions (Fig. 2). To further decipher the profiles of cleavage that can be obtained on the four nucleotides, we performed hierarchical clustering of the profiles and draw the corresponding heatmap (Fig. 3A). This allowed determining nine clusters of cleavage profiles. At first, we observed that in contrast to Thymine, the Adenine was the less stringent nucleotide to be accommodated by ncRVDs. We also noted that only two clusters contained clones presenting a nuclease activity almost restricted one nucleotide, with members of cluster 7 targeting nearly exclusively a Guanine and members of cluster 2 targeting predominantly an Adenine (Fig. 3A, Supplementary Fig. 6). All other clusters were targeting at least two nucleotides.

We further noticed that the most often occurring natural RVDs are not always found in the most specific clusters. For instance, NI, that is commonly used to target an Adenine, belonged to the first cluster (with relative activities: A/100%, C/73%, G/58%, T/0%) while the similar mutant MI is located in cluster 2 and turned out to be more specific (with relative activities: A/100%, C/0%, G/0%, T/0%).

We then determined the frequency of each amino-acid in each cluster in order to find out enrichment for particular amino-acids (Fig. 3b and Supplementary Fig. 6). We found, as expected, that the position 12, mainly involved in the stabilization of the RVD, was much more variable in each cluster than position 13 that presented important biases in some of the clusters. For instance, cluster 5 that is composed of clones targeting mainly an Adenine and a Guanine, contained predominantly the amino acids K, R, N and Q at position 13. Clones from cluster 1 that contains HD but also NI are targeting



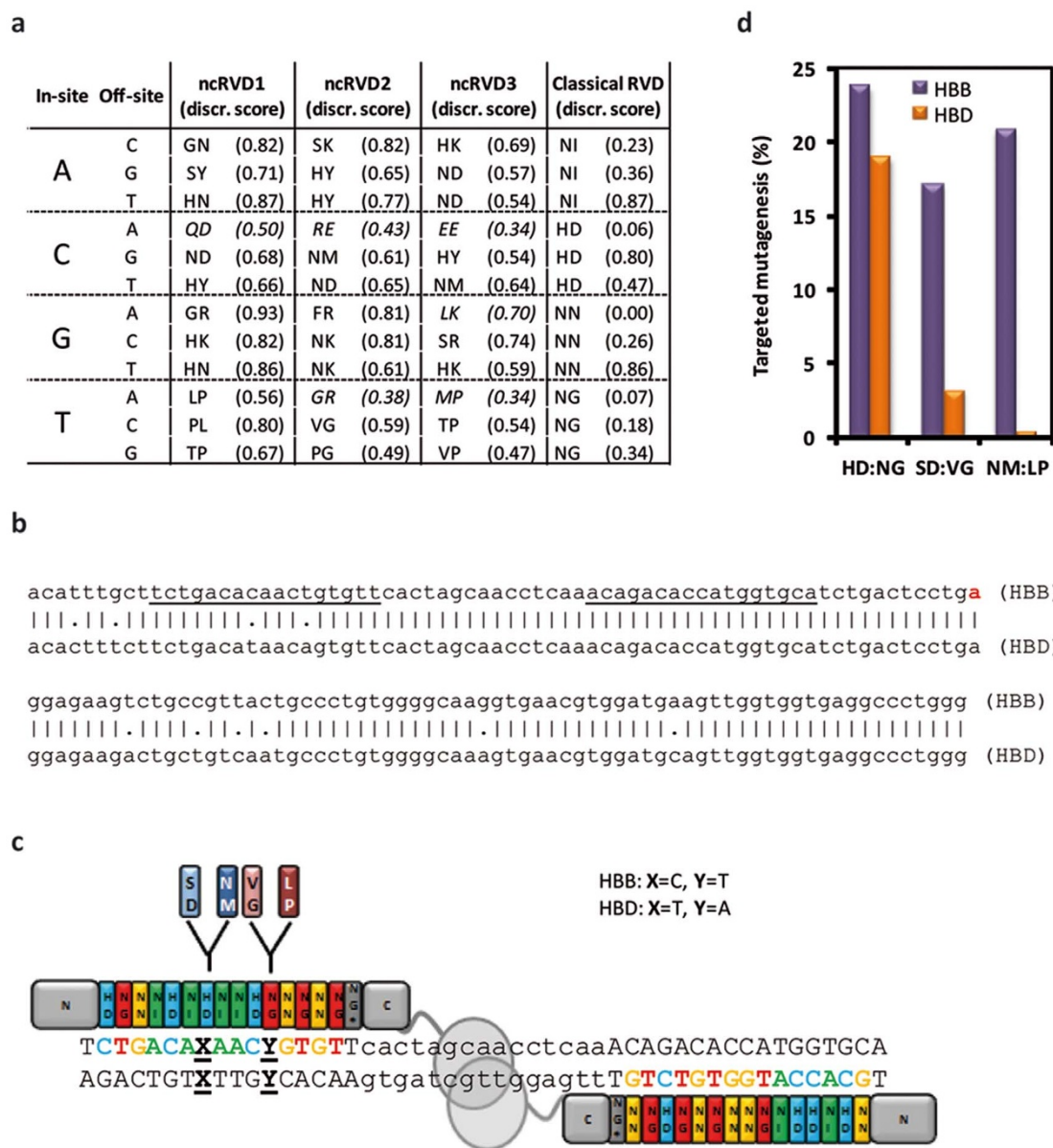
**Figure 3** | Hierarchical clustering of the cleavage profiles of ncRVDs on the 4 nucleotides. (a) Clusters are visualized by the alternation of red and blue and are numbered from top to bottom. Position of classical RVDs and ncRVDs used in the mammalian discrimination experiments are indicated. (b) Occurrence of every amino acid in clusters 1, 2 and 5. Each amino acid at position 12 and 13 is represented by a different color. The size of the cluster (number of different RVDs) and the targeting preference are indicated.

mainly an Adenine and a Cytosine and are enriched in acidic residues at position 13. Finally, cluster 2 which predominantly targets an Adenine is enriched in hydrophobic residues, essentially F and L (Fig. 3b).

**Use of ncRVDs to improve targeting specificity in vivo.** The ability to master the targeting specificity of TALEN by implementation of ncRVDs would be of great interest, notably to minimize off-site targeting. In contrast to previous work aiming to fully replace an RVD type in a TALE array<sup>16</sup>, we thought to develop an exclusion strategy focused on educated incorporation of limited number of ncRVD based on their discriminative properties. Based on our dataset, we thus selected ncRVDs that maintained a robust activity on a nucleotide while allowing discrimination of one or more other nucleotides when incorporated in a TALEN (Fig. 4a and Supplementary Fig. 7). As proof of concept, we aimed at targeting a sequence (in-site) close to the mutation responsible for the sickle cell disease in the HBB gene, knowing that this locus of interest is highly similar to a portion (off-site) of the HBD gene (93% of identity on the 140 base pairs surrounding the mutation, Fig. 4b).

Using features from our TALEN scaffold (TAL DNA binding array of 15.5 RVDs and spacer length of 15 bp), we found an optimal target that do not present within the human genome any off-site sequences with less than 5 mismatches at the exception of the HBD locus that contain 2 mismatches. The two HBB and HBD loci presenting 94% of identity with 32/34 identical nucleotides in the targeted sequence (Fig. 4b). The corresponding TALEN may therefore present significant processing of not only the HBB but also the HBD locus<sup>12,17,18</sup>. A careful examination of (i) the identity of the mismatches between the HBB and HBD sequences and (ii) our ncRVD dataset, enabled determining ncRVDs that would permit controlled targeting, ultimately excluding the processing of the HBD locus. We thus implemented either SD and VG or NM and LP as alternative ncRVDs to replace an HD and NG RVD in the classical TALEN (Fig. 4b and c).

To assess the discriminative potential of ncRVDs we performed targeted mutagenesis experiments at the HBB and HBD loci in a human cell line using one classical TALEN and two that contained the combination of ncRVDs (Fig. 4c). We monitored mutagenic events (small deletions and insertions, Indels), generated by the



**Figure 4 | Evaluation of discriminative properties on ncRVDs in mammalian cells.** (a) Potential discriminating ncRVDs identified in the library screens. These ncRVDs were determined especially by maximizing the difference of nuclease activity on two nucleotides (favouring a strong nuclease activity on the desired base and a weak or no nuclease activity on the other base). Italic indicates either non negligible residual activity on the weaker nucleotide, or an intermediate activity on the stronger nucleotide. Value in parenthesis represents the discrimination score (scale from 0, no discrimination to 1, maximum discrimination, arbitrary units, computed as the difference of intensities on both nucleotides). (b) Sequence alignment of the  $2 \times 69$  base pairs of HBB and HBD loci surrounding the position of the mutation responsible for the sickle cell anaemia. The two TALEN binding sequences are underlined. (c) Schematic representation of the loci and TALEN used in the mammalian cell target discrimination experiment. (d) Targeted mutagenesis frequency detected at the two endogenous loci (desired locus or in-site, HBB in violet and off-site, HBD in orange) using the HD/NG, SD/VG and NM/LP containing TALEN as measured by deep sequencing.

non-homologous end joining (NHEJ) repair pathway, at the double strand break (DSB) site induced by the TALEN. We thus quantified, by deep sequencing, targeted mutagenesis frequencies on specific PCR products surrounding the HBB or HBD loci. We observed that the three TALEN induced similar levels of targeted mutagenesis at the expected HBB locus (Table 1, Fig. 4d). Remarkably, while the classical TALEN was unable to discriminate between the two loci, the incorporation of the two NM and LP ncRVDs allowed a dramatic decreased of off-site targeting at the HBD locus (Fig. 4d).

Although a larger *in vivo* data set would be desirable to precisely define the limits of ncRVD incorporation in TALEN, taken together

our data indicates that implementation of such ncRVDs in TALENs is a valuable strategy to control and tune TALEN specificity.

## Discussion

The TALE scaffold has proven over the past few years to be one of the most robust platforms to generate molecular tools to edit, regulate and detect genes of interest<sup>19–30</sup>. However, as with all engineered DNA binding domains, depending on the experimental conditions and the targeted sequence, a certain extent of unwanted (off-site) targeting events might need to be controlled.

In this work, we aimed at extending the level of TALEN targeting control by taking advantage of the vast repertoire of possibilities



Table 1 | Activities of TALEN on their endogenous cognate target (HBB) and off-target sequence (HBD)

	locus	Targeted mutagenesis [%]	Events	WT	Reads
<b>HD/NG</b>	HBB	23,8	1200	3841	5041
	HBD	19,1	734	3120	3854
<b>NM/LP</b>	HBB	20,9	681	2578	3259
	HBD	0,4	46	11120	11166
<b>SD/VG</b>	HBB	17,2	515	2479	2994
	HBD	3,3	253	7462	7715

offered by the DNA targeting characteristics of RVDs (position 12 and 13). We and others recently demonstrated that only a small number of mismatches relative to the “standard” association code (NI:A, HD:C, NN:G and NG:T) are tolerated in a TALEN<sup>12,17,18</sup>. We thus reasoned that the use of ncRVDs would mainly benefit for TALEN presenting off-site targets with few mismatches (1–3 mismatches, representing approximately a fourth of all possible TALEN<sup>12</sup>) and that a complete replacement of one or several of the four more common natural RVDs would not be required. We were comforted in this strategy as previous studies reported mitigated results with complete replacement of classical RVDs<sup>16,31,32</sup>. We therefore proposed a strategy focused on the educated incorporation of a limited number of ncRVDs based on their base preference properties. In particular, their exclusion capacities that prevent targeting and processing of sequences containing defined nucleotides at the adequate position, expands the utility of this technology. We thus provide a panel of characterized ncRVDs that will allow researchers to easily implement this methodology, which is also compatible with published state-of-the-art TALE assembly protocols<sup>33–35</sup>. Using such approach, we were able to significantly improve the targeting specificity (reduce the off-site targeting) while maintaining the robustness (high level of gene targeting) of the TALE platform. Such improvement in the fine tuning of the targeting specificity using ncRVDs associated with other recent strategies<sup>17,18</sup> (e.g. reduction of RVD number, engineering of the TALE C-terminal domain) should allow approaching the single nucleotide discriminative resolution. We thus anticipate that ncRVDs can greatly benefit to nuclease design, especially for therapeutic applications and synthetic biology<sup>36</sup>.

## Methods

**Preparation of libraries and extrachromosomal SSA assay in yeast.** The library of degenerated RVDs randomized at amino acid positions 12 and 13 was created by short overlap PCR (SOE) using the conventional NNK codon degeneracy (encoding the complete set of natural amino acids) with the following oligonucleotides: Fragment 1A: CCCAGTCACGACGTTGTAAAAC/Fragment 1B: GTCTCC-AGCGCTGCTTGCCGCMNMMNGCTGGCGATGGCCACCACCTGCTC and Fragment 2A: CCCAGTCACGACGTTGTAAAAC/Fragment 2B: CACAGAAACAGCTATGACCATG. The DNA template consisted in a single tandem repeat commercially synthesized and subcloned in the pAPG10 plasmid (Top Gene Technologies)<sup>14</sup>. The second (assembly) PCR was realized using the following oligonucleotides: Assembly1: Biotin-CCCAGTCACGACGTTGTAAAAC/Assembly2: CCCGGTACCGATCTCGAGG. The resulting PCR product was purified and analysed by DNA sequencing to ensure the randomization of the two positions. This single randomized tandem repeat was used in the preparation of the four TALE array libraries. Assembly of the complete TALE arrays using solid support (magnetic beads) was performed as described in Daboussi et al.<sup>14</sup>. All intermediate arrays released from the solid support were subcloned in the pAPG10 plasmid, transformed in XL1b (Stratagene) according to standard molecular biology procedures. A number of colonies at least 5-fold larger than the theoretical library size (1024) was scraped from the agarose plate and DNA extracted according to standard procedures. The intermediate array was prepared for the next assembly step by performing enzymatic restriction using SfiI and BbvI as previously described<sup>14</sup>. The final arrays were subcloned in plasmids already containing a TALEN scaffold (described in refs 12, 37). The resulting constructs were reanalysed by DNA sequencing to ensure the randomization of the two positions.

For each of the four TALE array libraries, plasmidic DNA of 1152 clones were extracted and transformed in yeast cells with a success rate >91% (colony growth). The nuclease activity screens in yeast were performed for each of the yeast clones as previously described in Juillerat et al.<sup>12</sup>.

For experiments in the mammalian cell line (Human 293H cells, Life Technologies), all TALE arrays were obtained from Collectis Bioresearch (Paris, France) and the TALE-based nucleases were expressed under the control of an EF1a promoter (described in ref. 37). TALEN<sup>TM</sup> is a trademark owned by Collectis Bioresearch.

**Nuclease transfection and targeted mutagenesis.** Cell culture and nuclease transfections were performed as described in Juillerat et al.<sup>12</sup>. Cells were pelleted by centrifugation and genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer’s instructions. Targeted mutagenesis was estimated on the whole cell population three days post transfection. PCR of the endogenous loci (410 bp for HBB, In-site and 400 bp for HBD, Off-site) were performed using the following oligonucleotides: In-site1: CCACACCCTAGGG-TTGCCCAATCTACTCCC/In-site2: GTAGACCACCAGCAGCCTAAGGGTGGG and Off-site1: GCCTCCTTGAGCCTCTTTATAAC/Off-site2: GCGGTGGGGA-GATATGTAGAGGAG. PCR products were purified using the AMPure kit (Invitrogen). Amplicons were further analyzed by deep-sequencing<sup>38</sup>.

**Bioinformatics analysis.** The distance profile was computed as follows: for all identical RVDs found active in different collections, we identified all pairs with one member in one collection and the other member in another collection, each member having its own nuclease activity profile on the four nucleotides. We then computed the Euclidean distance between these two profiles as a measure of similarity (the lower, the most similar), giving a distribution of distances. As a point of comparison, we computed the same distance distribution between pairs of different RVDs from the same collection. Smoothed lines were computed as kernel density estimates with a Gaussian kernel and a bandwidth of 0.1.

Clustering analysis was done using hierarchical clustering utilizing the Euclidean distance between profiles and complete linkage method.

Discriminating ncRVDs were identified as follows: for each nucleotide pair N1 and N2 for which we want to find a ncRVD that cuts N1 but not N2, the solutions (ncRVDs 1 to 3) were first sorted by occurrence number (number of collections in which the RVD has been observed to cut N1), and then by the difference of nuclease activity between N1 and N2. In particular, we filtered out ncRVDs having an intensity on N1 smaller than 0.75 and an intensity on N2 larger than 0.30 (nuclease activities, arbitrary units). Remaining ncRVD candidates were ordered by decreasing value difference between N1 and N2. The first three best RVDs for (N1, N2) were retained. If there were less than 3 solutions, these solutions were completed to 3 by adding all ncRVDs having a positive difference between the intensities on N1 and N2 without taking into consideration the occurrence number and the constraints limits (simple ordering by decreasing difference).

All computations were done using the R statistical software package (<http://cran.r-project.org/index.html>).

- Scharenberg, A. M., Duchateau, P. & Smith, J. Genome engineering with TALE-effector nucleases and alternative modular nuclease technologies. *Curr Gene Ther* **13**, 291–303 (2013).
- Gersbach, C. A. & Perez-Pinera, P. Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opin Ther Targets* **18**, 835–839 (2014).
- Niu, J., Zhang, B. & Chen, H. Applications of TALENs and CRISPR/Cas9 in human cells and their potentials for gene therapy. *Mol Biotechnol* **56**, 681–688 (2014).
- Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
- Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
- Deng, D. et al. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720–723 (2012).
- Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716–719 (2012).
- Stella, S. et al. Structure of the AvrB3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallogr D Biol Crystallogr* **69**, 1707–1716 (2013).
- Cong, L., Zhou, R., Kuo, Y. C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968 (2012).



10. Streubel, J., Blucher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol* **30**, 593–595 (2012).
11. Daboussi, F. *et al.* Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Res* **40**, 6367–6379 (2012).
12. Juillerat, A. *et al.* Comprehensive analysis of the specificity of transcription activator-like effector nucleases. *Nucleic Acids Res* **42**, 5390–5402 (2014).
13. Arnould, S. *et al.* Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* **355**, 443–458 (2006).
14. Daboussi, F. *et al.* Genome engineering empowers the diatom *Phaeodactylum tricorutum* for biotechnology. *Nat Commun* **5**, 3831 (2014).
15. Meckler, J. F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118–4128 (2013).
16. Yang, J. *et al.* Complete decoding of TAL effectors for DNA recognition. *Cell Res* **24**, 628–631 (2014).
17. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833–838 (2013).
18. Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **11**, 429–435 (2014).
19. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149–153 (2011).
20. Mercer, A. C., Gaj, T., Fuller, R. P. & Barbas, C. F. 3rd Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res* **40**, 11163–11172 (2012).
21. Mercer, A. C., Gaj, T., Sirk, S. J., Lamb, B. M. & Barbas, C. F. 3rd Regulation of Endogenous Human Gene Expression by Ligand-Inducible TALE Transcription Factors. *ACS Synth Biol* **3**, 723–730 (2013).
22. Yanik, M. *et al.* TALE-PvuII Fusion Proteins - Novel Tools for Gene Targeting. *PLoS One* **8**, e82539 (2013).
23. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137–1142 (2013).
24. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M. E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* **20**, 1321–1324 (2013).
25. Beurdeley, M. *et al.* Compact designer TALENs for efficient genome engineering. *Nat Commun* **4**, 1762 (2013).
26. Boissel, S. *et al.* megaTALS: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res* **42**, 2591–2601 (2014).
27. Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* **31**, 1133–1136 (2013).
28. Owens, J. B. *et al.* Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. *Nucleic Acids Res* **41**, 9197–9207 (2013).
29. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
30. Sun, N. & Zhao, H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng* **110**, 1811–1821 (2013).
31. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* **29**, 699–700 (2011).
32. Christian, M. L. *et al.* Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One* **7**, e45383 (2012).
33. Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Honing, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* **31**, 76–81 (2013).
34. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460–465 (2012).
35. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).
36. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* **40**, 7584–7595 (2012).
37. Juillerat, A. *et al.* Exploring the transcription activator-like effectors scaffold versatility to expand the toolbox of designer nucleases. *BMC Mol Biol* **15**, 13 (2014).
38. Valton, J. *et al.* Overcoming TALE DNA Binding Domain Sensitivity to Cytosine Methylation. *J Biol Chem* **287**, 38427–38432 (2012).

## Author contributions

A.J. and P.D. conceived the study and designed experiments. A.J., C.P., G.D. and V.G. performed experiments. A.M., J.V., F.D. and L.P. provided conceptual advices. A.J. and A.D. analyzed experiments. A.J., A.D. and P.D. wrote the manuscript with support from all authors.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** All co-authors are present or former Collectis employees.

**How to cite this article:** Juillerat, A. *et al.* Optimized tuning of TALEN specificity using non-conventional RVDs. *Sci. Rep.* **5**, 8150; DOI:10.1038/srep08150 (2015).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>