Supporting Information

Engineering a Robust DNA Split Proximity Circuit with Minimized Circuit Leakage

Yan Shan Ang, Rachel Tong and Lin-Yue Lanry Yung*

Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585

Table of Content

S1. List of DNA sequences and domain designs

Table S1. List of all DNA sequences used in this study. All DNA sequences were designed using Nupack web server to minimise secondary structure and misfoldings.(1) The domains were colour-coded according to Fig. 1 in the main text.

Strand	Sequence
Trigger Strand	
Direct Trigger (**DT**)	GCC CTT ACT CCC AAT TCC AAC
Hairpin Trigger (**HT**)	GCC CTT ACT CCC GCG GCG CGC CTT GGC GCG CCG CTT AAT TCC AAC
Split Trigger (**ST**)	GAG TGG ATG GTG AAG GTG AAG GTA
Initiators Design Rules	
Initiator 2 (**I2-3'**)	GCC CTT ACT CC_ GGT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC
Initiator 2 (**I2-5'**)	_CC CTT ACT CCC GGT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC
Initiator 1-9 nt Split (**I1-S9**)	TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG ACC TTA ATT CCA AC
Initiator 1-10 nt Split (**I1-S10**)	TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG ACC TTC AAT TCC AAC
Initiator 2-Design 1 (**I2-D1**)	GCC CTT ACT CCC GGT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC
Initiator 2-Design 2-Lock (**I2-D2L**)	GGG AGT AAG GGC TTT
Initiator 2-Design 3 (**I2-D3**)	GCC CTT ACT CCG GGT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC
Initiator 2-Design 4 (**I2-D4**)	GCC CTT ACT CC_ GGT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC
Study of Intermediate Complexes Responsible for Background Leakage	
Initiator 1 (I1-*length of domain a-length of domain b*)	**I1-a3-b9**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG TT C AAT TCC AAC **I1-a4-b9**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG C TT C AAT TCC AAC **I1-a4-b6**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG C TT C AAT TCC **I1-a5-b9**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG AC TT C AAT TCC AAC **I1-a6-b9**: Refer to I1-S10 **I1-a6-b7**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG ACC TT C AAT TCC A **I1-a6-b5**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG ACC

	TT C AAT TC **I1-a6-b4**: TAC CTT CAC CTT TTT TTT TTT TTT TTT GTG ACC TT C AAT T
Initiator 2 (I2-*length of domain a*)	**I2-a3**: GCC CTT ACT CC CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC **I2-a4**: GCC CTT ACT CC G CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC **I2-a5**: GCC CTT ACT CC GT CAC TTT TTT TTT TTT TTT CAC CAT CCA CTC **I2-a6**: Refer to I2-D4
Streptavidin-Biotin as Model System	
Streptavidin-Initiator 1 (**Stav-I1**)	[1] /5-Biotin/TTT TTT TTT TTT TTT GTG CTT CAA TTC CAA C
Streptavidin-Initiator 2 (**Stav-I2**)	GCC CTT ACT CCG CAC TTT TTT TTT TTT TTT /3-Biotin/ [1]
	Thrombin Binding Aptamers as Recognition Moiety
I1-fibrinogen binding exosite (**Throm-I1**)	GGT TGG TGT GGT TGG TTT TTT TTT GTG CTT CAA TTC CAA C
I2-heparin binding exosite (**Throm-I2**)	GCC CTT ACT CCG CAC TTT TTT TTT TTT TTT AGT CCG TGG TAG GGC AGG TTG GGG TGA CT
Reporter Signals	
HCR Hairpin 1 (**HP1**)	GTT GGA ATT GGG AGT AAG GGC TGT GAT GCC CTT ACT CCC
HCR Hairpin 2 (**HP2**)	GCC CTT ACT CCC AAT TCC AAC GGG AGT AAG GGC ATC ACA
Fluorescein-tagged readout (**F**)	GTT GGA ATT GGG AGT AAG GGC /3-FAM/ [1]
Dabcyl-Tagged Quencher (**Q**)	[1] /5-Q/ GCC CTT ACT CCC

[1] Special modification to the DNA termini are denoted by the notation /"5' or 3' end"-"type of modification"/

Table S2. List of domain sequences involved in the DNA circuit design.

Domain	Description	Length (nt)	Sequence
a	Association region	3	GTG
		4	GTG C
		5	GTG AC
		6	GTG ACC
b	Reporter toehold	6	GGA ATT
		7	T GGA ATT
		8	TT GGA ATT
		9	GTT GGA ATT
c	Readout branch migration site	12	GGG AGT AAG GGC
d	Alternating toehold of HCR	6	TGT GAT
s	Distance modulator (or "spacer")	15	TTT TTT TTT TTT TTT
x	**ST** recognition region 1	12	GAG TGG ATG GTG
y	**ST** recognition region 2	12	AAG GTG AAG GTA
apt1	Aptamer for fibrinogen binding exosite	15	GGT TGG TGT GGT TGG
apt2	Aptamer for heparin binding exosite	29	AGT CCG TGG TAG GGC AGG TTG GGG TGA CT

S2. Further Characterization of Leak I

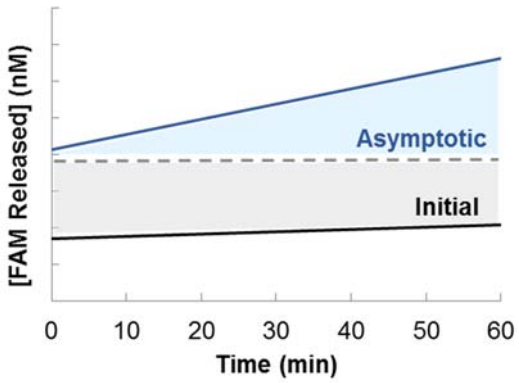

Figure S1. Leak I could be further decoupled into types of leakages: (1) initial due to oligonucleotide synthesis defect (grey region), and (2) asymptotic defect due to spurious hybridization (blue region). The former was characterized by an initial signal burst while the latter was characterized by a steady growth of noise over time.

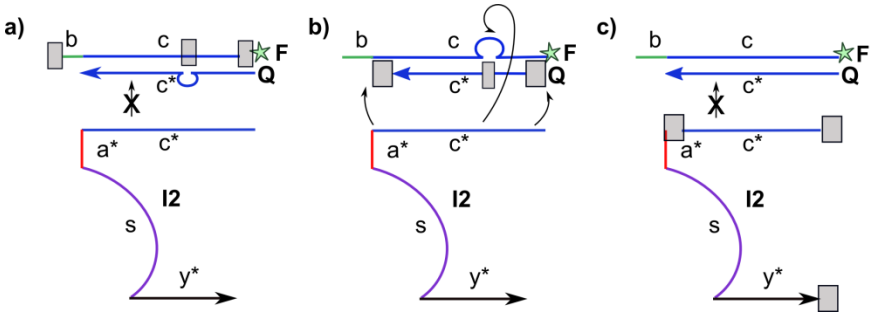

Figure S2. Three sites of synthesis defect (3', 5' and internal), as indicated by grey boxes, are possible for the n-1 oligonucleotide product of a) fluorophore (**F**) strand, b) quencher (**Q**) strand and c) initiator 2 (**I2**) strand. By logical deduction, the presence of defects on **F** and **I2** cannot release the reporter signal. On the other hand, defects on **Q** can expose sites on the **F-Q** complex for strand displacement by **I2** to form a more thermodynamically-stable **I2-F** product. In our work, however, we did not consider defects in internal sites due to the inherent randomness in their position.

S3. Other Designs to Suppress Asymptotic Leakage in Leak I

We conceptualized three other designs to combat the asymptotic leakage in Design 1 (Fig. S3a). In Design 2, domain *c** of **I2** was protected with a lock strand (**I2-D2L**). The next two designs focused on single nucleotide changes at the 3'-end of domain *c**, i.e. single mismatch (Design 3) or translocation to the 5'-end of domain *b** in **I1** (Design 4).

All three designs (Designs 2 – 4) successfully suppressed the asymptotic leakage as seen from the relatively stable background noise over time (Fig. S3b). Design 2 was most effective in reducing the initial leakage (87.1% reduction in initial fluorescent burst) compared to the 13.5% and 16.6% reduction by Design 3 and 4 respectively. This was because the nucleotide at the 3'-end of domain *c** in Design 2 was completely locked up by the complementary **I2-D2L** strand. On the other hand, the same nucleotide was only made energetically less accessible by introducing the single mismatch in Design 3. The leakage in Design 4 was comparable to that in Design 3 as the next nucleotide on the 5'-end of domain *a** was also guanine and so both designs effectively presented the same sequence for the toehold-independent strand displacement of **Q** strand in a **F-Q + I2** reaction mixture.

However, the positive signal generated in presence of **ST** was heavily penalized in Design 2 due to the unfavourable strand displacement kinetics across a four-way junction (Fig. S3c) (2). For Design 3, the single mismatch introduced at the first incumbent nucleotide of domain *c** hindered the effective initiation of branch migration. Since branch migration is a nucleotide-by-nucleotide random walk process, the first nucleotide after the toehold domain plays a crucial role in determining whether the branch migration continues forward or the bound toehold duplex dissociates (3).

These observations were confirmed using gel electrophoresis in junction with HCR readout to amplify the minute amount of circuit leakage (Fig. S3d).

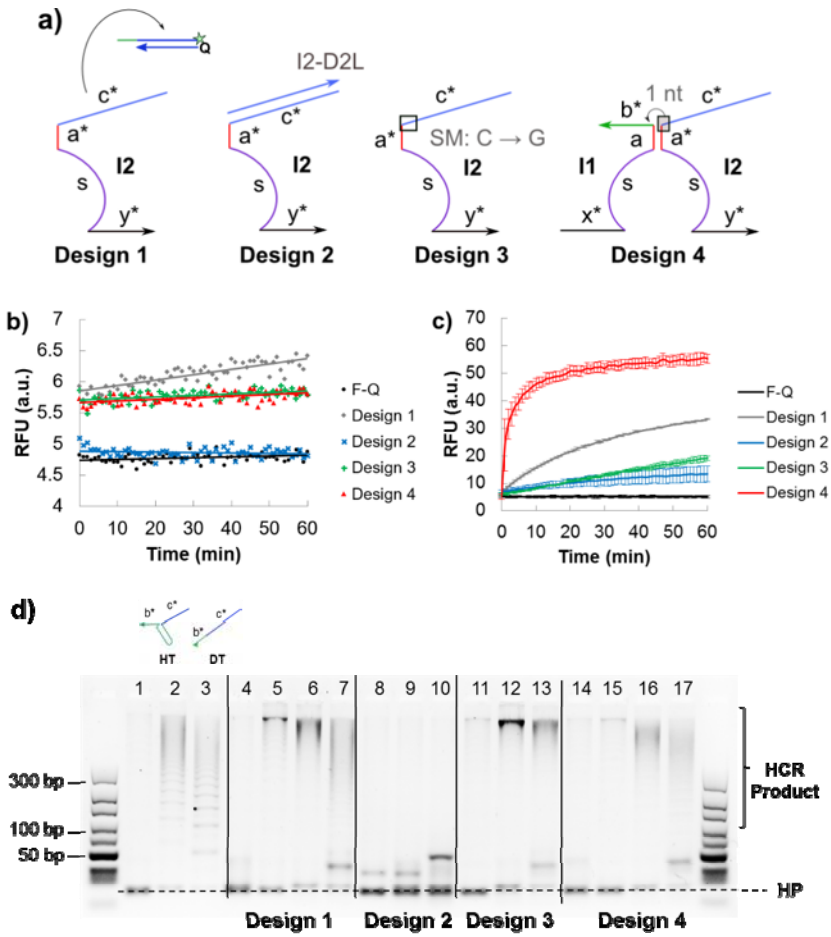

Figure S3. (a) Three strategies (Designs 2 – 4) were conceptualized to suppress the toehold-independent strand displacement (Design 1). (b) The generation of Type I leakage over time was evaluated in a **F-Q + I2** reaction mixture. The mean values of triplicate readings were presented as scatter plot, while error bars were not shown for the narrow RFU range to avoid congesting the plot. (c) Time-profile of positive signals generated in the presence of 100 nM **ST** and all circuit components (**F-Q + I1 + I2**) shows that Design 4 generated the highest signal intensity at the fastest rate. Error bars shown indicate sample standard deviation of triplicate experiments. All reaction mixtures contained 100 nM of the relevant DNA components. (d) Gel electrophoresis was used to qualitatively evaluate the **I2** designs. Lane 1 corresponds to hairpins (**HP1** and **HP2**) only, lanes 2 and 3 correspond to positive controls using hairpin trigger (**HT**) and direct trigger (**DT**) respectively. Lanes representing each design are demarcated by solid lines and are shown in the following sequence (from left to right): **HP1** + **HP2 + I2** (lanes 5, 8, 11 and 15)**, HP1** + **HP2** + **I1** + **I2** (lanes 6, 9, 12 and 16)**,** and **HP1** + **HP2** + **I1** + **I2** + target strand (lanes 7, 10, 13 and 17). Additional lanes of **HP1** + **HP2** + **I1** are shown for Designs 1 (lane 4) and 4 (lane 14). 500 nM of individual DNA components was used. 10 – 300 bp DNA ladders are shown at both sides of the gel.

S4. Further Characterization of Leak II

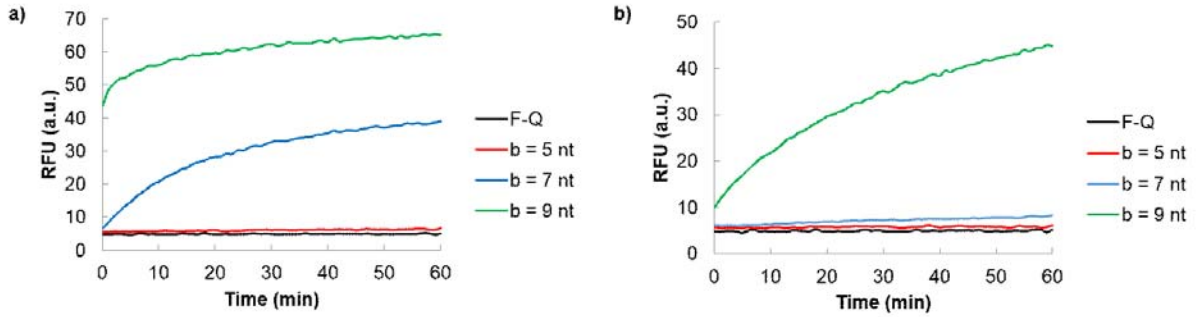

Figure S4. a) Positive signal (**F-Q** + **I1** + **I2** + **ST)** and b) background noise (**F-Q** + **I1** + **I2)** at various toehold lengths (length of domain *b** = 5, 7 and 9 nt). The length of association region (domain *a**) was kept constant at 6 nt. The concentration of all DNA components (**F-Q, I1, I2** and **ST**), added whenever relevant, was 100 nM.

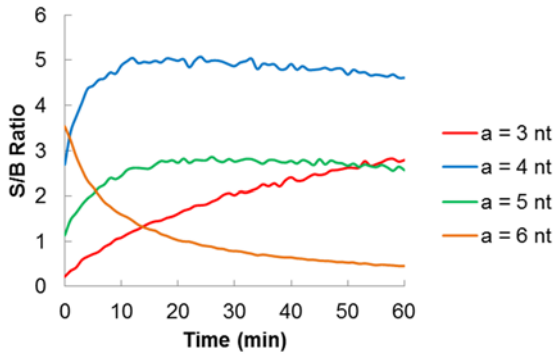

Figure S5. Temporal evolution of the signal-to-background (S/B) ratio at various association lengths (domain *a** = 3, 4, 5 and 6 nt). The toehold length (domain *b**) was kept constant at 9 nt. The concentration of all DNA components (**F-Q, I1, I2** and **ST**) was 100 nM.

S5. Role of domain *b** as the de facto reporter toehold

We argue that that domain *b**, which is the actual toehold designed to trigger the reporter signal, should be kept at a kinetically functional length (≥ 6 nt) to achieve decent signal-to-background (S/B) ratio (4). The underlying assumption is that while strand displacement can be triggered by dynamic toehold binding (a non-equilibrium event), the intermediate complexes receiving the incoming strand should be sufficiently stable at equilibrium. This is based on the understanding that a simultaneous three-entity collision leading to meaningful product formation is an extremely rare event.

We consider the case where the length of domain *b** is greater than that of domain *a**, which is now constrained to ≤ 5 nt (Figure S6a). Under this constraint, no **I1-I2** complex forms at equilibrium for domain *a** < 5 nt under the reaction condition specified in this work according to Nupack simulation (1). In the presence of the split target (**ST**) strand, the local concentrations of **I1** and **I2** increase significantly resulting in the higher probability of them associating momentarily at domain *a** to form a high level of the proper **ST-I1-I2** assembly (major product). The trade-off is that a stable intermediate complex can form between **I1** and **F-Q** at domain *b** (minor product) from which domain *a** now serves as the new reporter trigger toehold. However, due to the short length of domain *a**, the kinetics of strand displacement is slow and equilibrium yield is low.

Conversely, the length of domain *b** can be shortened such it is shorter than domain *a** (Figure S6b). This disrupts the formation of any stable **I1-F-Q** complex. However, due to the short length of domain *b**, strand displacement of the reporter signal is kinetically slow. At the same time, a relatively significant amount of stable **I1-I2** complex are present at equilibrium. This implies that the effect of additional association due to the increased local concentrations of initiator strands in the presence of **ST** is less pronounced.

From the above analysis, we infer that having a longer domain *b** length than domain *a** length is preferred under the constraint that domain *a** should be kept sufficiently short, i.e. <5 nt, such that **I1-I2** complex does not exist at equilibrium. We further support our logical argument using an experimental example involving two cases where the lengths of domains *a** and *b** were interchanged between 4 nt and 6 nt (Figure S6c). When the length of domain *b** was longer than domain *a** (as in Figure S6a), discernible level of signal was generated over the background. Conversely, when the length of domain *a** was longer than domain *b** (as in Figure S6b), both signal and background were generated at a similarly low level. These

experimental observations were in line with our expectations stated above. Hence, both the logical argument and experimental results reinforced the importance of keeping the de facto toehold, domain *b**, at a sufficiently long length (≥ 6 nt) to ensure good signal-to-noise ratio.

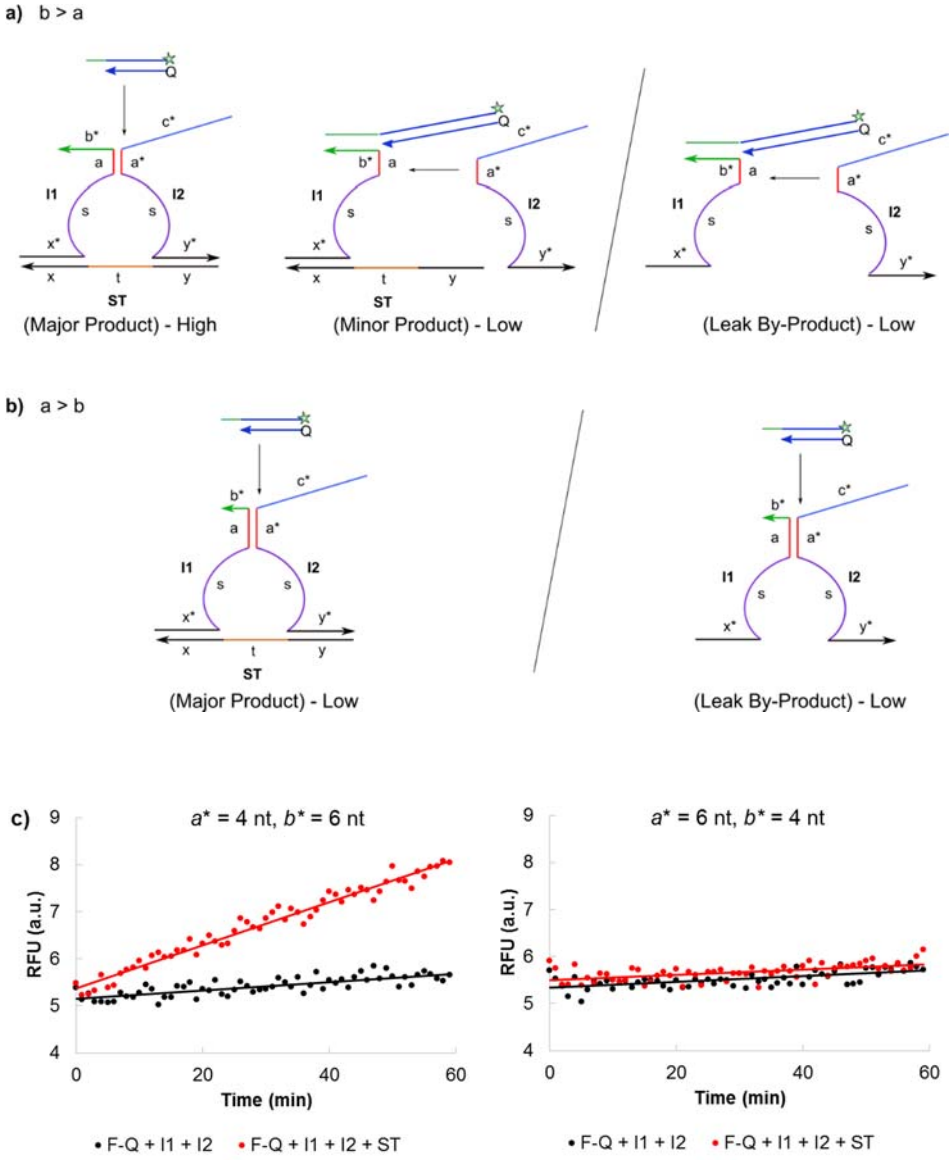

Figure S6. Identification of major and minor products as well as leak by-products generated when a) length of domain *b** is greater than domain *a*;* b) length of domain *a** is greater than domain *b**. The terms "high" and "low" indicate the expected level of complex formation relative to the total amount of input DNA strands. The diagonal line partitions the generation of signal (left) and noise (right). c) We illustrate the above analysis using an example involving two cases, i.e. (*a** = 4 nt, *b** = 6 nt) and (*a** = 6 nt, *b** = 4 nt). The former configuration resulted in appreciable signal generated in presence of **ST** whereas for the latter case, both signal and background were generated at a similarly low levels such that they could not be differentiated.

S6. Characterization of the Final DNA Split Proximity Assay

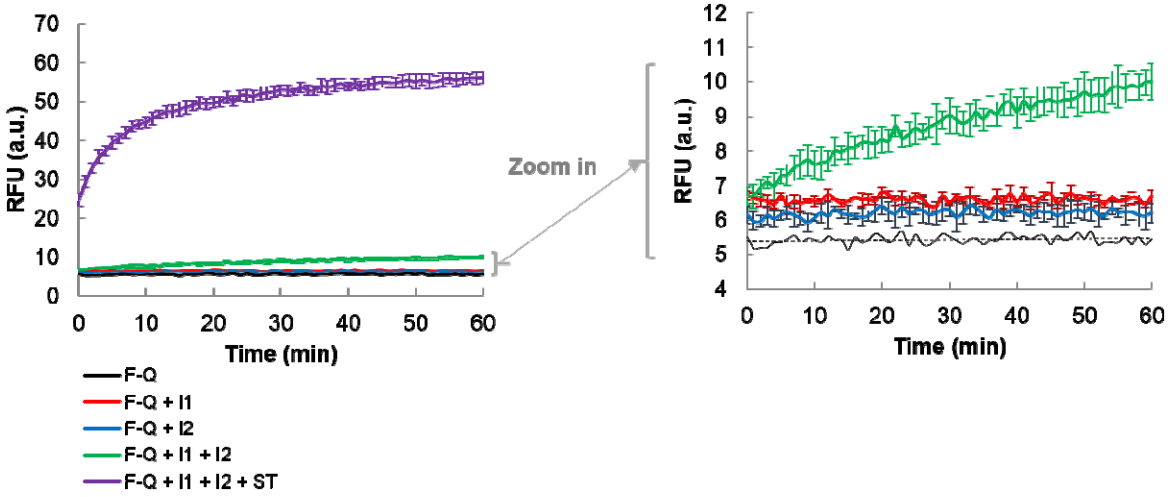

Figure S7. The final DNA split proximity assay was designed with domain *a** = 4 nt and domain *b** = 9 nt. The generation of positive signal was improved while minimizing the circuit leakage. Unfortunately, Leak II was contributed by an intermediate complex essential for the generation of positive signals, i.e. **I1-F-Q**, and could not be completely eliminated. The concentration of all DNA components (**F-Q, I1, I2** and **ST**), added whenever relevant, was 100 nM.

References

1. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170-173.
2. Khodakov, D.A., Khodakova, A.S., Huang, D.M., Linacre, A. and Ellis, A.V. (2015) Protected DNA strand displacement for enhanced single nucleotide discrimination in double-stranded DNA. *Sci. Rep.*, **5**, 8721.
3. Srinivas, N., Ouldridge, T.E., Šulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.*, **41**, 8886-8895.
4. Zhang, D.Y. and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.*, **3**, 103-113.