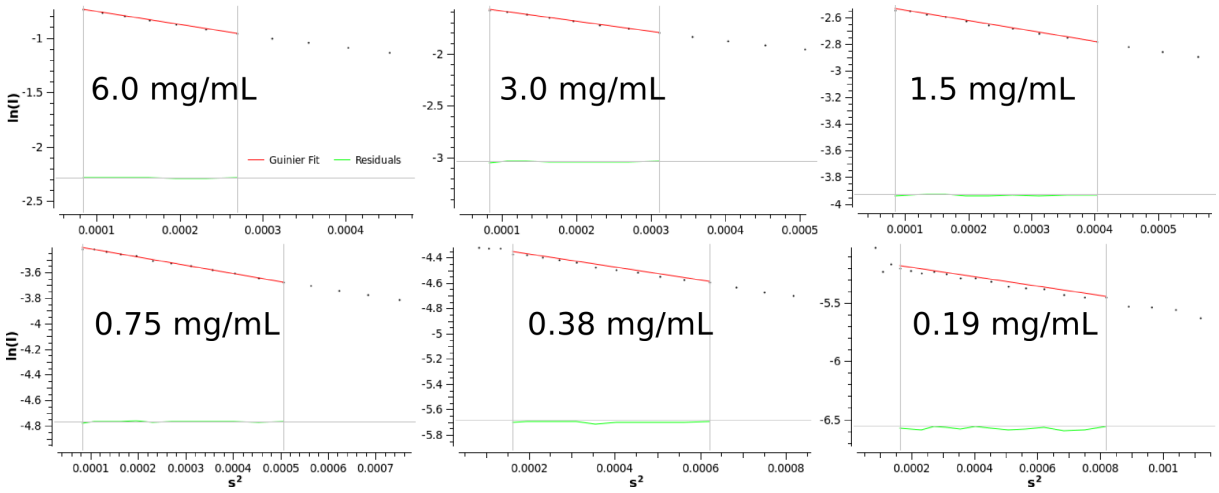
Supporting Information

SI Figures – SI Tables

The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation

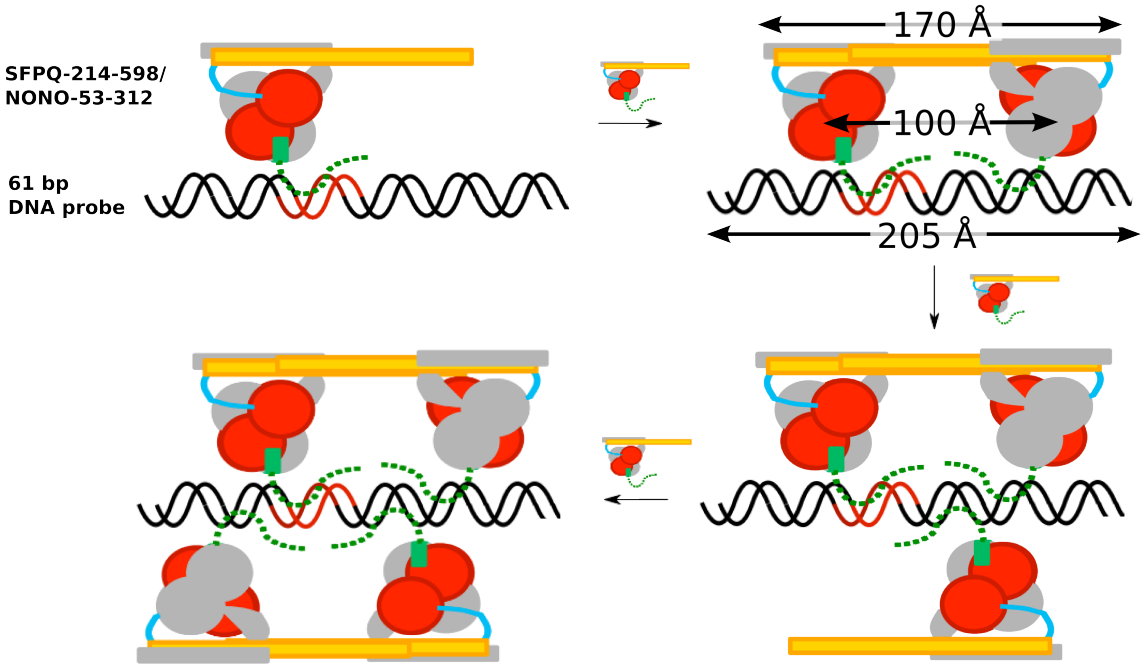
Mihwa Lee, Agata Sadowska, Indra Bekere, Diwei Ho, Benjamin Gully, Yanling Lu, K. Swaminathan Iyer, Jill Trewhella, Archa H. Fox* and Charles S. Bond*

Supplementary Figure 1



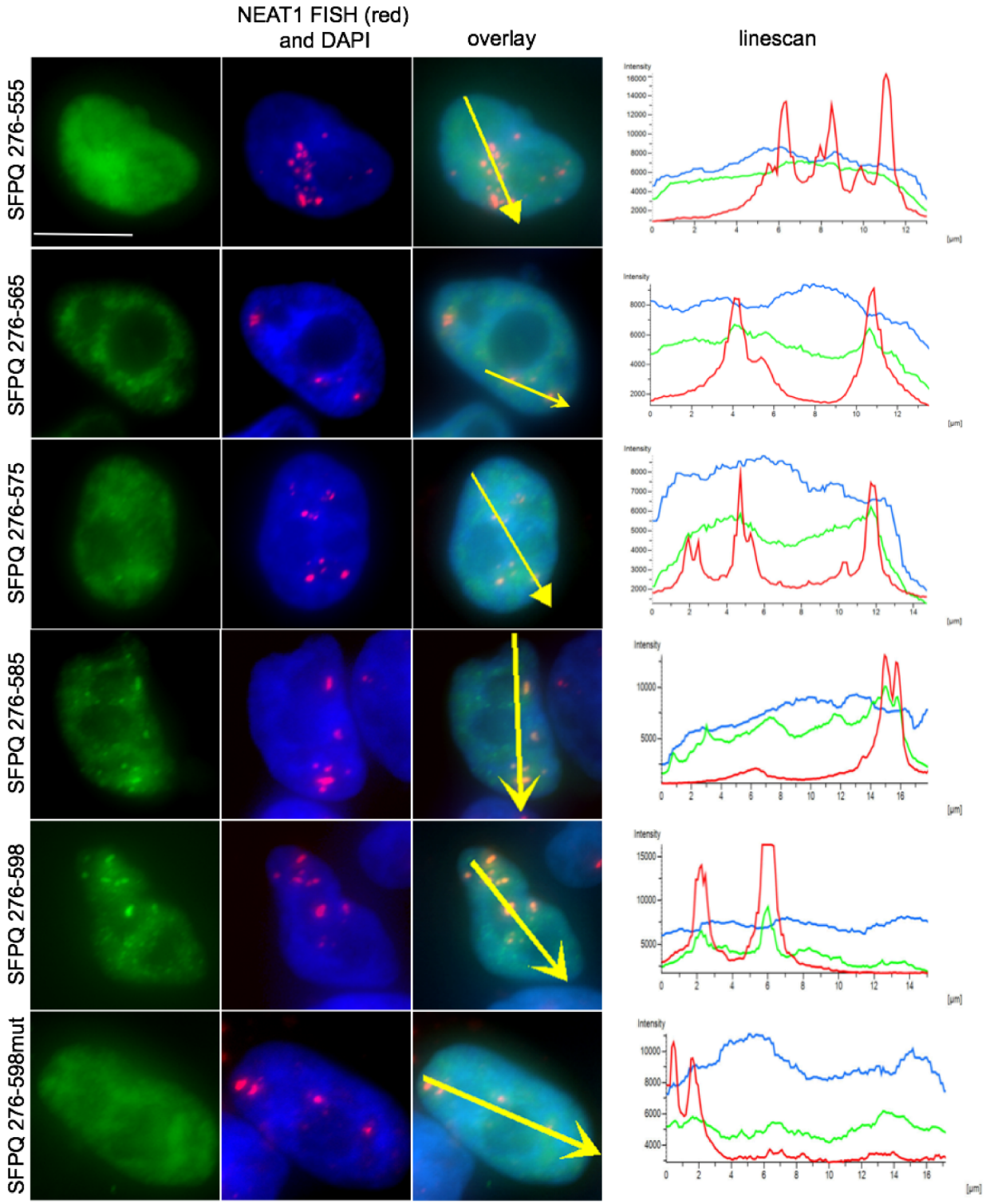
Guinier analysis for Small angle X-ray scattering data of a dilution series (6.0 – 0.19 mg/mL) of the SFPQ-276-598/NONO-53-312 heterodimer (see **Fig 3**).

Supplementary Figure 2



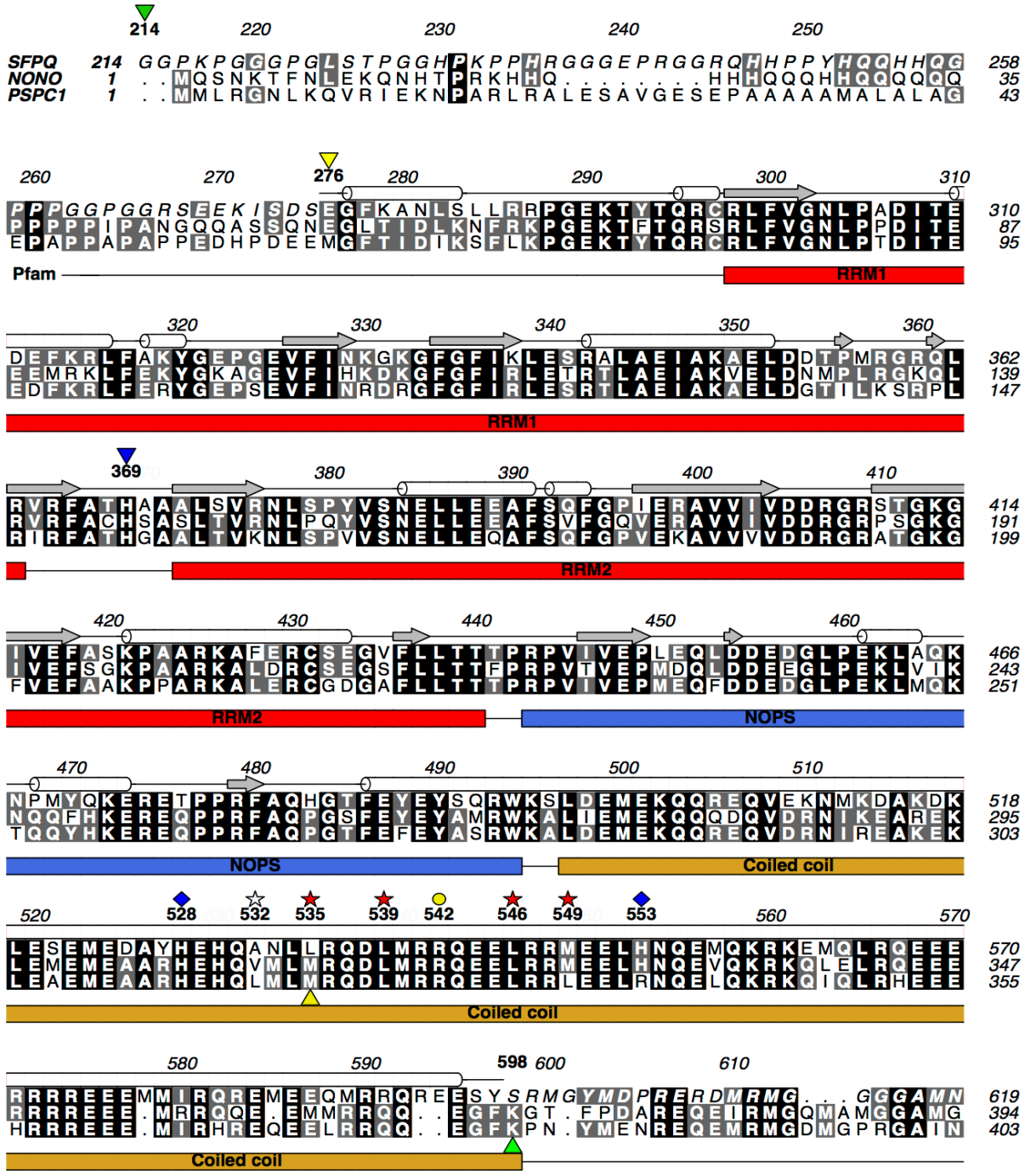
Proposed DNA binding model in EMSA (**Fig 4**). Two RRMs, NOPS domain, and coiled-coil domain in SFPQ are depicted as red circles, cyan line, and yellow rectangle, respectively. Equivalent domains of NONO are presented in grey. Solid green block represents the N-terminal short α-helical structure observed in SFPQ-276-535. Residues 214-275 (not resolved in crystal structures) are depicted in green dashed line. The dimensions of the GAGE6 probe used would match a maximum of two heterotetramers binding per probe.

Supplementary Figure 3



SFPQ positive foci are paraspeckles that colocalize with NEAT1 RNA. For each GFP-fused truncated SFPQ protein (corresponding to proteins "1"-"5" in **Fig 4**) fluorescence microscopy data is given (scale bar 5 micron). Left, GFP-SFPQ fusion (green); center NEAT1 FISH (red) and DAPI (blue); right, overlay. Yellow vectors drawn on the overlay correspond to linescans presented on the right (green, red and blue corresponding to GFP, NEAT1 and DAPI). Colocalization is identifiable by coincident peaks for GFP and NEAT1, as exemplified in SFPQ-276-598.

Supplementary Figure 4



Sequence alignment of the conserved DBHS domain of the three human DBHS proteins. A representative secondary structure scheme deduced from three crystal structures of SFPQ reported in this study is drawn on top of the sequences. Domain structure defined in PFAM is displayed under the sequences. Boundaries for the constructs reported in this study are indicated by the triangles with down-pointing triangles for N-terminus and the upward-pointing ones for C-terminus. Residues involved in the coiled-coil interaction motif are marked with diamonds, stars and a circle.

Supplementary Table 1

Structural Parameters from SAXS Data and Dimer Model Fit Results

Protein conc* (mg/mL)	Guinier analysis ($q \cdot R_g < 1.0$**)			$P(r)$ analysis			Dimer model fit
	R_g (Å)	$I(0)$ x 10^{10} cm	$I(0)/c$***	R_g (Å)	$I(0)$ x 10^{10} cm	d_{max} (Å)	χ
6.0	60 ± 0.6	0.5286 ± 0.0023	0.0881	61.1 ± 0.2	0.5247 ± 0.0012	230	n.a.
3.0	54 ± 0.6	0.2246 ± 0.0011	0.0749	55.8 ± 0.2	0.2242 ± 0.0005	210	n.a.
1.5	48 ± 0.7	0.0846 ± 0.0005	0.0564	48.9 ± 0.2	0.0838 ± 0.0003	170	5.7
0.75	44 ± 1.0	0.0350 ± 0.0004	0.0467	45.9 ± 0.6	0.0352 ± 0.0002	170	2.0
0.38	40 ± 1.7	0.0139 ± 0.0003	0.0371	41.6 ± 0.8	0.0140 ± 0.0002	150	0.7
0.19	35 ± 2.8	0.0059 ± 0.0002	0.0315	38.6 ± 1.2	0.0061 ± 0.0001	140	0.4

* Protein stock solution of was determined as 6.0 mg/mL by absorption at 280 nm. Subsequent serial two-fold dilutions were used to prepare the concentration series.

** Guinier analysis was confined to $q \cdot R_g < 1.0$ due to the highly asymmetric shapes of the scattering particles.

*** For a stable single species, $I(0)/c$ would be constant with increasing c. When c is expressed in mg/mL there is a linear relationship between the molecular weight of the scattering particle and $I(0)/c$, thus if $I(0)/c = 0.0371$ is the dimer, then a tetramer would give $I(0)/c = 0.0742$.

n.a. is not applicable