

 Open access • Posted Content • DOI:10.1101/2021.10.08.463684

Poly-glutamine-dependent self-association as a potential mechanism for regulation of androgen receptor activity — [Source link](#)

Josep Rizo, Carlos M. Roggero, Victoria Esser, Lingling Duan ...+5 more authors

Institutions: University of Texas Southwestern Medical Center

Published on: 08 Oct 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Androgen receptor and Dihydrotestosterone

Related papers:

- [Sintokamide A Is a Novel Antagonist of Androgen Receptor That Uniquely Binds Activation Function-1 in Its Amino-terminal Domain.](#)
- [Nucleolin represses transcription of the androgen receptor gene through a G-quadruplex.](#)
- [Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells \(steroid receptor associated protein/dihydrotestosterone\)](#)
- [Dual androgen receptor \(AR\) and STAT3 inhibition by a compound targeting the AR amino-terminal domain](#)
- [Functional Motifs of the Androgen Receptor](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/poly-glutamine-dependent-self-association-as-a-potential-2sw0kli8e4>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

**Poly-glutamine-dependent self-association as a potential mechanism
for regulation of androgen receptor activity**

Carlos M. Roggero^{1,2,3*}, Victoria Esser^{1,2,3}, Lingling Duan⁴, Allyson M. Rice¹, Shihong Ma⁶, Ganesh V. Raj⁶,
Michael K. Rosen^{1,7}, Zhi-Ping Liu^{4,5} and Josep Rizo^{1,2,3*}

¹Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, United States;
²Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas, United
States; ³Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas,
United States; ⁴Department of Internal Medicine, University of Texas Southwestern Medical Center,
Dallas, Texas, United States; ⁵Department of Molecular Biology, University of Texas Southwestern Medical
Center, Dallas, Texas, United States; ⁶Department of Urology, University of Texas Southwestern Medical
Center, Dallas, Texas, United States; ⁷Howard Hughes Medical Institute, University of Texas Southwestern
Medical Center, Dallas, Texas, United States

*Email: Carlos.Roggero@UTSouthwestern.edu or Jose.Rizo-Rey@UTSouthwestern.edu

Short title: Regulation of androgen receptor activity by self-association

24 **Abstract**

25 **The androgen receptor (AR) plays a central role in prostate cancer. Development of castration resistant**
26 **prostate cancer (CRPC) requires androgen-independent activation of AR, which involves its large N-**
27 **terminal domain (NTD) and entails dramatic epigenetic changes depending in part on histone lysine**
28 **demethylases (KDMs) that interact with AR. The AR-NTD is rich in low-complexity sequences, including**
29 **a polyQ repeat. Longer polyQ sequences were reported to decrease transcriptional activity and to**
30 **protect against prostate cancer. However, the molecular mechanisms underlying these observations**
31 **are unclear. Using NMR spectroscopy, here we identify weak interactions between the AR-NTD and the**
32 **KDM4A catalytic domain, and between the AR ligand-binding domain and a central KDM4A region that**
33 **also contains low-complexity sequences. We also show that the AR-NTD can undergo liquid-liquid phase**
34 **separation in vitro, with longer polyQ sequences phase separating more readily. Moreover, longer**
35 **polyQ sequences hinder nuclear localization in the absence of hormone and increase the propensity for**
36 **formation of AR-containing puncta in the nucleus of cells treated with dihydrotestosterone. These**
37 **results lead us to hypothesize that polyQ-dependent liquid-liquid phase separation may provide a**
38 **mechanism to decrease the transcriptional activity of AR, potentially opening new opportunities to**
39 **design effective therapies against CRPC.**

40

41

42

43

44 **Introduction**

45 Prostate cancer is one of the most common forms of cancer in men(1). Survival and proliferation of
46 prostate cancer cells depend critically on signaling through the androgen receptor (AR) and, consequently,
47 standard therapies against prostate cancer involve androgen reduction through chemical or surgical
48 castration(2, 3). Unfortunately, prostate cancer cells eventually develop the ability to activate AR
49 independently of androgen and patients invariably develop the more aggressive castration-resistant
50 prostate cancer (CRPC), which is associated with a much worse prognosis(4-6). Hence, understanding the
51 mechanisms underlying AR function and how AR is activated in a ligand-independent manner is crucial to
52 develop effective therapies against CRPC.

53 One aspect of AR biology that is most likely involved in prostate cancer is the complicated
54 functional interplay between AR and epigenetic enzymes, including histone deacetylases (HDACs) and
55 lysine demethylases (KDMs). Thus, AR activity can be downregulated or upregulated by HDACs and KDMs
56 (e.g. (7-11)), and the development and progression of prostate cancer is accompanied by extensive
57 abnormalities in the levels of these enzymes and in histone modification marks(12, 13), which has been
58 referred to as an ‘epigenetic catastrophe’(14). Moreover, AR was reported to bind to several KDMs,
59 including KDM4A, although only limited information was reported about the regions of the KDMs and of
60 AR implicated in the interactions(9-11). Hence, the nature of these interactions remains unclear and it is
61 unknown how they affect AR activity.

62 AR contains a long N-terminal domain (NTD), a central DNA-binding domain (DBD), a short hinge
63 region (H) and a ligand-binding domain (LBD) (Fig. 1A). The LBD is normally key for receptor activation,
64 but the NTD is required for full transcriptional activity, and interactions between NTD and LBD regulate
65 androgen-dependent gene expression(15-17). Activation involves the translocation of AR to the nucleus
66 when androgen binding to the LBD induces a large conformational change that exposes a nuclear
67 localization signal (NLS) present between the DBD and LBD(18). The NTD has attracted much attention

68 not only because of its known participation in transcriptional activity but also, and in particular, because
69 diverse C-terminally truncated splice variants of AR lacking the LBD have been identified in prostate cancer
70 cell lines and in clinical specimens(19). These variants can explain the development of androgen-
71 independent AR activity in CRPC. Indeed, multiple studies have suggested that AR-V7, which is the most
72 abundant of these variants and is constitutively localized to the nucleus, mediates androgen-independent
73 cell growth and resistance to androgen deprivation (reviewed in ref. (20)). Interestingly, our recent finding
74 that KDM4B regulates the generation of AR-V7 by alternative splicing upon androgen deprivation(21)
75 provided another connection between AR and KDM function.

76 The AR NTD contains abundant low-complexity sequences that are common in transactivation
77 domains of transcription factors and that include polyQ and poly-G regions with variable length in the
78 human population(22). Exceedingly long polyQ sequences (> 38 Qs) can lead to neurodegenerative
79 disease(23), which may be associated to the formation of oligomeric fibrils(24). Conversely, short polyQ
80 length (< 20 Qs) was shown to increase the risk for prostate cancer(25, 26), indicating that longer polyQ
81 sequences may have a protective effect. These findings may arise because the transcriptional potency of
82 AR is inversely correlated with polyQ length(27-29). As expected, the NTD is largely unstructured in
83 solution(30), but some sequences do have defined propensities to form secondary structure, particularly
84 in regions that are important for transactivation(31, 32). The polyQ sequence has a clear tendency to form
85 α -helical structure that increases with the length of the sequence and is stabilized by interactions of the
86 glutamine side chains with the backbone(33-35). These observations led to the proposal that the effects
87 of polyQ length on transcriptional activity may arise from changes in the strength of protein-protein
88 interactions, for instance those with transcriptional co-regulators or general transcription factors(34). It
89 is also important to note that the AR NTD has been shown to undergo liquid-liquid phase separation at
90 high concentrations (100 μ M)(36), and that a propensity for phase separation appears to be a
91 fundamental property of many low-complexity sequences that are involved in transcriptional activation

92 and/or mediate formation of membraneless organelles(37-39). However, the importance of this property
93 for AR function and the influence of polyQ length on the tendency of the AR NTD to undergo phase
94 separation have not been explored.

95 The study presented here was designed to examine how AR binds to KDM4A and to investigate
96 how self-association of the AR NTD and nuclear localization of AR depend on the length of the polyQ
97 sequence. NMR experiments reveal weak interactions between the AR-NTD and the KDM4A catalytic
98 domain, and between the AR LBD and a central KDM4A region that also contains low-complexity
99 sequences. We also show that the drive for the AR-NTD to undergo phase separation in vitro increases
100 progressively with polyQ length. We verify that polyQ length is inversely correlated with transcriptional
101 activity in two prostate cancer cell lines. Moreover, we find that nuclear localization is decreased by longer
102 polyQ sequences in the absence of hormone, and that such sequences increase the formation of AR-
103 containing puncta in nuclei of cells treated with dihydrotestosterone (DHT). Based on these results, we
104 propose that longer polyQ sequences in AR may have a protective role against prostate cancer by
105 enhancing self-association processes that reduce the AR transcriptional activity.

106

107 **Results**

108 **Direct AR-KDM4A binding through regions containing low-complexity sequences**

109 The present study was initiated with the goal of gaining insights into how KDMs affect AR function by
110 investigating the nature of the interactions between these proteins. KDM4A and its homologues contain
111 an N-terminal catalytic domain, a central region that contains low-complexity sequences, a C-terminal
112 region that includes two PHD domains and two Tudor domains (Fig. 1A). Co-immunoprecipitation
113 experiments suggested that AR binds to KDM4A and KDM4D through its LBD, while the catalytic domain
114 and the C-terminus of KDM4A, or the C-terminus of KDM4D, mediate AR binding(11). However, it is
115 difficult to derive definitive conclusions with regard to direct physical interactions from co-

116 immunoprecipitation. To examine whether we could observe direct binding between purified
117 recombinant proteins, we performed extensive pulldown assays and hold up assays(40) using fragments
118 that spanned different regions of AR and KDM4A. However, we were unable to obtain definitive results
119 using these approaches, suggesting that potential interactions between AR and KDM4A are too weak to
120 be reliably detected by these methods.

121 We turned to an NMR method based on the analysis of perturbations caused by an unlabeled
122 protein on transverse relaxation optimized (TROSY) ^1H - ^{15}N heteronuclear single quantum coherence
123 (HSQC) spectra of a uniformly ^{15}N -labeled protein, which provide a sensitive tool to detect interactions
124 between the two proteins(41). In a first set of experiments, we acquired ^1H - ^{15}N TROSY-HSQC spectra of a
125 ^{15}N -labeled fragment spanning the catalytic domain of KDM4A (residues 1-350) in the absence and
126 presence of different unlabeled fragments of AR. DHT was included in experiments with fragments
127 containing the AR LBD, which is unstable without a ligand. A fragment spanning the NTD of AR (AR-NTD;
128 residues 1-559) caused noticeable but limited broadening on the ^1H - ^{15}N TROSY-HSQC spectrum of ^{15}N -
129 KDM4A(1-350), while a longer fragment that included the AR NTD and DBD (AR-NTD-DBD; residues 1-632)
130 caused stronger broadening (Fig. 1B,C). In contrast, almost no perturbations on the ^1H - ^{15}N TROSY-HSQC
131 spectrum of ^{15}N -KDM4A(1-350) were caused by fragments containing the AR LBD alone (AR-LBD; residues
132 663-919) or from the DBD to the LBD (AR-DBD-LBD; residues 554-919) (Fig. 1D,E). These results show that,
133 under these conditions, the catalytic domain of KDM4A binds weakly to AR-NTD and that such binding is
134 enhanced by the DBD, while there is no appreciable binding to the LBD or to the DBD in the absence of
135 the NTD. Hence, binding of the AR-NTD-DBD fragment to KDM4A(1-350) appears to be driven by weak
136 interactions with the NTD that can cooperate with additional interactions involving the DBD. The latter
137 might be too weak to be observable without such cooperation or might be hindered by intramolecular
138 binding of the DBD to the hinge region or the LBD in the context of the AR-DBD-LBD fragment.

139 Because the AR-LBD had been previously implicated in binding to KDM4s(11) but we did not
140 observe binding to KDM4A(1-350), we tested whether the LBD might interact with other regions of
141 KDM4A. For this purpose, we prepared ¹⁵N-labeled AR-LBD bound to DHT and acquired ¹H-¹⁵N TROSY-
142 HSQC spectra in the absence and presence of fragments spanning the central region of KDM4A (residues
143 301-708) or its C-terminal region containing the PHD and Tudor domains (residues 703-1064). The central
144 region of KDM4A containing the low complexity sequences caused substantial broadening on the ¹H-¹⁵N
145 TROSY-HSQC spectrum of ¹⁵N-AR-LBD, indicative of binding, while the C-terminal KDM4A fragment
146 induced very little perturbations (Fig. 1F,G).

147 We also analyzed whether the AR-NTD-DBD fragment binds to the catalytic domains of two other
148 KDM4 isoforms, KDM4B and KDM4C. The ¹H-¹⁵N TROSY-HSQC spectrum of ¹⁵N-labeled catalytic domain of
149 KDM4C exhibited dramatic broadening upon addition of AR-NTD-DBD, while the broadening was less
150 overt in the ¹H-¹⁵N TROSY-HSQC spectrum of ¹⁵N-labeled catalytic domain of KDM4B (Supporting Fig. S1).
151 Hence, there is some level of specificity in the interaction of AR-NTD-DBD with KDM4 isoforms.

152 To confirm the interaction of AR-NTD-DBD with the catalytic domain of KDM4A by another
153 method, we performed chemical cross-linking experiments with bis(sulfosuccinimidyl)suberate (BS3), an
154 agent that cross-links primary amines with primary amines. When we treated mixtures of AR-NTD-DBD
155 and KDM4A(1-350) with BS3 and analyzed the mixture by SDS-PAGE followed by coomassie blue staining,
156 we observed the almost complete disappearance of the bands of AR-NTD-DBD and KDM4A(1-350) and
157 the appearance of bands corresponding to high molecular weight cross-linked complexes (Fig. 2A). An AR
158 fragment spanning most of the NTD but lacking the DBD (residues 93-495) was also cross-linked efficiently
159 with KDM4A(1-350) (Fig. 2B), albeit not as quantitatively as the AR-NTD-DBD fragment. Hence, these data
160 correlate with the NMR results, which revealed binding of AR-NTD to KDM4A(1-350) that was
161 strengthened by inclusion of the AR-DBD (Fig. 1B,C).

162 Overall, our results indicate that there are indeed direct interactions between AR and KDM4A.
163 The interactions appear to be weak, but may cooperate with other interactions among components of
164 the transcriptional machinery for recruitment of these proteins and/or control of their activities.
165 Interestingly, the two types of interactions that we observed involve regions containing low complexity
166 sequences of AR (binding of AR-NTD to the KDM4A catalytic domain) or KDM4A (binding of the KDM4A
167 central region to AR-LBD).

168

169 **Longer polyQ sequences enhance gel formation by the AR-NTD**

170 Increasing experimental evidence shows that many proteins that are multivalent for protein-protein
171 interactions have a tendency to undergo liquid-liquid phase separation and that this property underlies
172 the formation of membraneless subcellular compartments that serve a wide variety of biological
173 functions(38). In such systems, the drive to phase separate increases with protein length and/or
174 interaction valence(38, 42). The tendency to phase separate is particularly common for low-complexity
175 sequences present in transcription factors and other proteins involved in formation of complexes with
176 nucleic acids (e.g. (37, 39)). Indeed, the AR-NTD (residues 1-559) was shown to phase separate into liquid
177 droplets at 100 μ M concentration(36). These findings, together with the known importance of the AR-
178 NTD for development of CRPC, the notion that longer polyQ sequences in the AR-NTD may protect against
179 prostate cancer, and our finding that low complexity sequences mediate AR-KDM4A interactions, led us
180 to investigate how polyQ length affects the ability of AR-NTD to phase separate. For this purpose, we
181 prepared and purified recombinant fragments corresponding to the AR-NTD that contained polyQ
182 sequences with 12, 20, 31 and 49 glutamines (AR-NTD-12Q, -20Q, -31Q and -49Q, respectively).

183 The four versions of AR-NTD each exhibited some turbidity at 500-600 μ M concentrations and
184 200 mM NaCl, consistent with the previous observation of phase separation by this AR fragment(36), but
185 the turbidity was markedly stronger for the AR-NTD fragments with longer polyQ sequence (Fig. 3A). After

186 a sample of AR-NTD-31Q was allowed to reach high turbidity, the sample became clear upon addition of
187 10% 1,6-hexanediol (Fig. 3B). This observation shows that phase separation by NTD is reversible and does
188 not result from irreversible formation of amyloid aggregates(43). To have a more quantitative measure of
189 the influence of polyQ length on the drive to phase separate, we prepared different dilutions of the four
190 AR-NTD fragments from freshly prepared concentrated solutions, incubated the resulting samples for 1 h
191 at 4 °C, room temperature (RT; ca. 23 °C) or 37 °C, and measured the absorption at 400 nm. Analysis by
192 SDS PAGE illustrated the purity of the samples used in these experiments (Supporting Fig. S2).
193 Precipitation prevented reliable quantification of the turbidity at 37 °C, but consistent results were
194 obtained in the experiments performed at 4 °C and RT, which generally revealed progressive increases of
195 absorption at 400 nm with the concentration for the four proteins. Importantly, the turbidity correlated
196 with the length of the polyQ sequence under the majority of conditions (Fig. 3C,D; note that the lower OD
197 at 400 nm observed for 100 μM AR-NTD-49Q after 1 h at RT, compared to that observed for 50 μM AR-
198 NTD-49Q or 100 μM AR-NTD-31Q, might arise because droplets may settle out of solution). These results
199 indicate that longer polyQ sequences increase the drive for the AR-NTD to phase separate.

200 The length of the polyQ sequence in AR was found to be inversely correlated with transcriptional
201 activity(27-29). To verify these findings with the same polyQ tracts that we used for the turbidity assays,
202 we used a luciferase assay in two prostate cancer cell lines (LNCaP and PC3) transfected with vectors
203 expressing full-length AR that contained 12, 20, 31 and 49 glutamines (AR-12Q, -20Q, -31Q and -49Q,
204 respectively). In LNCaP cells, which express endogenous AR, we observed robust enhancement of
205 transcriptional activity upon addition of DHT even when cells were transfected with control vector (Fig.
206 4A). The DHT-induced increase in activity was markedly stronger for cells transfected with AR-12Q, but
207 the stimulation gradually decreased as the polyQ length increased, and there was no additional
208 stimulation in cells expressing AR-49Q (Fig. 4A). To ensure that these results did not arise from lower
209 expression of the AR variants with longer polyQ sequences, we analyzed the total AR levels in Western

210 blots with AR antibodies and the levels of transfected AR proteins, which included a FLAG tag, with FLAG
211 antibodies (Fig. 4B). These experiments showed that the expression levels of the transfected AR proteins
212 were comparable to each other, and similar to the levels of endogenous AR. In experiments performed
213 with PC3 cells, which do not express endogenous AR, we did not observe a substantial DHT-induced
214 enhancement of transcriptional activity when the cells were transfected with control vector, but
215 transfection with AR-12Q caused robust DHT-dependent increase in activity and such increase again
216 decreased gradually with the length of the polyQ sequence (Fig. 4C). Hence, these results confirm that the
217 length of the polyQ sequence can indeed have strong effects on AR activity.

218

219 **Longer polyQ sequences decrease nuclear localization of AR and increase DHT-dependent formation of**
220 **AR-containing puncta in the nucleus**

221 Transcriptional activation by AR requires the transport of AR to the nucleus, which is normally induced by
222 ligand binding and exposure of a hidden NLS(18). However, C-terminally truncated AR variants such as AR-
223 V7 can be constitutively localized to the nucleus(20). To analyze how the length of the polyQ sequence
224 influences the nuclear localization of AR, we transfected LNCaP cells with Cherry-labeled AR variants
225 containing 12, 20, 31 and 49 glutamines, and analyzed their distribution in the cytoplasm and the nucleus.
226 The results of these experiments need to be examined with caution, as we observed nuclear localization
227 of the transfected proteins even in the absence of hormone (Fig. 5A,B), which may arise because
228 fluorescent proteins have a tendency to translocate to the nucleus(44). Nevertheless, these experiments
229 can still provide information regarding how the length of the polyQ sequence affects the propensity of AR
230 for nuclear localization. Importantly, we observed that, under castration-mimicking conditions lacking
231 hormone, longer polyQ sequences clearly favored the cytoplasmic localization of AR in detriment to its
232 nuclear localization (Fig. 5A,B). This finding did not arise from the differences in expression, as the levels
233 of the expressed cherry-AR proteins assessed by Western blot exhibited some variability (Fig. 5C) but this

234 variability did not correlate with the percentage of protein localized to the nucleus. In the presence of
235 DHT, most of the AR was localized to the nucleus regardless of polyQ length (Fig. 6A). However, we
236 observed that the nuclei contained puncta with higher Cherry fluorescence than the diffuse background,
237 and that longer polyQ sequences led to more, brighter and larger puncta (Fig. 6A). This conclusion was
238 confirmed by measuring the ratio of fluorescence intensity in puncta to that of the diffuse background,
239 which showed that the ratio correlated with polyQ length (Fig. 6B). These results suggest that there are
240 at least two potential mechanisms by which the increased tendency of AR to self-associate as polyQ length
241 increases may decrease its transcriptional activity, i.e. a decreased tendency to translocate to the nucleus
242 and an increased tendency to form nuclear puncta that might sequester AR and thus limit its access to the
243 transcriptional machinery.

244

245 **Discussion**

246 AR plays a key role in the development of CRPC. Understanding the mechanisms that control the
247 transcriptional activity of AR and that underlie the emergence of androgen-independent AR activity is thus
248 critical to design effective therapies to treat this devastating disease. The presence in AR of a large,
249 unstructured NTD rich in low complexity sequences, together with growing evidence that this type of
250 sequence often mediates incorporation of proteins into phase-separated membraneless organelles(38),
251 suggest that one function of the AR NTD may be to mediate phase separation or at least to enhance the
252 localization of AR into biomolecular condensates, thus modulating AR availability and/or activity. The
253 study presented here now shows that low-complexity regions of both AR and KDM4A are involved in
254 interactions between the two proteins in vitro and hence might contribute to epigenetic regulation of AR
255 activity. Moreover, we find that the length of the polyQ sequence correlates with the tendencies of AR-
256 NTD to phase separate in vitro, to remain in the cytoplasm and to form puncta inside the nucleus in the
257 presence of DHT. Altogether, these findings lead us to propose the hypothesis that polyQ-dependent

258 phase separation of AR may provide a means of negatively regulating its transcriptional activity. Other
259 explanations of our data are possible, however, and further studies will be needed to establish whether
260 the nuclear foci observed here indeed form by liquid-liquid phase separation in cells, and to learn whether
261 and how focus formation may be related mechanistically to effects on AR-mediated transcription.

262 Early work showed that the NTD is largely unstructured(30), but some of its sequences exhibit
263 clear propensities to adopt secondary structure(45). In particular, the polyQ sequence of the AR NTD has
264 an intrinsic tendency to form α -helical structure that increases in stability with the length of the
265 sequence(33-35). A leucine-rich region preceding the polyQ sequence also increases its propensity to form
266 α -helical structure and impairs aggregation(35). Conversely, exceedingly long polyQ sequences in AR favor
267 aggregation that most likely involves beta-amyloid structure and can lead to formation of toxic fibrils,
268 causing neurodegenerative disease(23, 24). Since longer polyQ sequences protect against prostate
269 cancer(25, 26) and lead to lower transcriptional activity(27-29) (Fig. 4), it appears that the optimal length
270 of the polyQ sequence may be the result of a trade-off between preventing over-activation of AR (if polyQ
271 is too short) and minimizing the possibility of aggregation that leads to neurodegeneration (if polyQ is too
272 long)(34, 46). The increased stability of the helical structure formed by longer polyQ sequences was
273 proposed to influence aggregation through formation of helical oligomers and to affect the transcriptional
274 activity of AR by altering the strength of protein-protein interactions(34). Our finding that the length of
275 the polyQ sequence correlates with the tendency of the AR NTD to phase separate (Fig. 3) shows that
276 longer polyQ sequences increase self-association. However, it is unclear whether such self-association
277 involves direct formation of oligomers with beta-structure or oligomerization via helix-helix interactions
278 that may or may not convert to beta-amyloid-like structures. It is even plausible that there are multiple
279 pathways to aggregation and the different pathways have different biological consequences (see below).
280 Regardless of these possibilities, the observation that addition of 1,6-hexanediol reverses gel formation

281 (Fig. 3B) indicates that the self-association of the AR-NTD is reversible and therefore does not involve the
282 irreversible formation of amyloid-like aggregates.

283 The correlation between polyQ length and propensity of the AR-NTD to phase separate (Fig. 3),
284 together with the inverse correlation between polyQ length and transcriptional activity(27-29) (Fig. 4),
285 suggest that polyQ-dependent phase separation of AR could provide a mechanism to reduce its activity.
286 However, it is still unclear how the activity is reduced, and whether reduction directly arises from
287 increased self-association. Phase separation can increase the efficiency of a biological process by helping
288 to concentrate multiple factors involved in the process in a defined compartment, but can also decrease
289 biological activity by sequestering a protein and thus limiting access to its site of action, or by actively
290 inhibiting it(47). Our analyses of nuclear versus cytoplasmic localization lead us to two hypotheses that
291 are not necessarily exclusive of each other. One hypothesis is based on the observation that longer polyQ
292 sequences hinder the nuclear localization of AR in the absence of hormone (Fig. 5), and postulates that
293 longer polyQ sequences enhance the formation of AR oligomers in the cytoplasm, hindering access to the
294 nuclear import machinery. Note that, although we did not observe formation of puncta containing AR in
295 the cytoplasm, oligomers of limited sizes would not be distinguishable in our experiments. The second
296 hypothesis arises from the observation that longer polyQ sequences led to increased formation of puncta
297 in the nucleus when DHT was present (Fig. 6), and proposes that localization of AR to these puncta
298 sequesters the protein and hence prevents its incorporation into transcriptional complexes. One
299 prediction of this hypothesis is that the AR foci do not correspond to AR-responsive genes, and that
300 localization of AR to its proper loci is decreased with increasing polyQ length.

301 It is important to realize that other mechanisms may also underlie the reduction of AR activity
302 caused by long polyQ sequences, such as increased affinity for transcriptional repressors. Moreover, it is
303 unclear whether AR can undergo phase separation by itself in a cellular environment, as high
304 concentrations of the AR-NTD are required for phase separation(36). It is also plausible that the AR NTD

305 facilitates recruitment to membraneless organelles that are formed through phase separation of other
306 proteins. Indeed, the AR NTD was readily incorporated in phase separated liquid droplets formed by
307 speckle-type POZ protein (SPOP), an adaptor for a ubiquitin ligase that helps to recruit substrates for
308 proteasomal degradation(36). Thus, increased recruitment of AR to subcellular compartments for protein
309 degradation might also underlie a decrease in AR activity with increased polyQ length. An additional
310 possibility is that the AR NTD mediates differential incorporation into more than one type of
311 membraneless organelle based on its length. For instance, AR might be recruited to organelles where it is
312 sequestered or degraded and to other organelles where transcription of AR-dependent genes is activated,
313 and where interactions between AR and KDM4s help recruiting either one of the two proteins to the
314 condensate. Longer polyQ sequences may tilt the balance in favor of recruitment to the former type of
315 condensate. Intriguingly, overexpression of the AR NTD was reported to delay progression of prostate
316 cancer tumors and CRPC(48, 49). The rationale behind these experiments was to use the AR NTD as a
317 decoy molecule that competes with endogenous AR for binding to proteins required for its activation, and
318 no interaction between the overexpressed AR NTD and endogenous AR was observed by co-
319 immunoprecipitation. However, it is plausible that weak interactions underlying phase separation might
320 not be detected by this method and that the decoy molecules helped to sequester endogenous AR
321 through phase separation. If this notion is correct, overexpression of AR NTD containing a moderately long
322 polyQ sequence (e.g. with 31 Qs) might provide a more effective means to delay progression of CRPC.

323 Clearly, there are many uncertainties about the ideas discussed above and much needs to be
324 learned to understand how AR becomes activated in an androgen-independent manner in CRPC. Our
325 results bring new hypotheses to this area that need to be tested with further research and that could open
326 new therapeutic avenues.

327

328 **Materials and methods**

329

330 **Protein expression and purification**

331 Standard recombinant DNA techniques were used to prepare pGEX-KG or pET28 expression vectors
332 encoding the AR and KDM4 fragments used in this study starting from vectors encoding the full-length
333 proteins. His₆-AR-NTD fragments (aa 1-559) containing 12Q, 20Q, 31Q or 49Q were expressed with pET28
334 vectors at 25 °C in *E. coli* BL21 (DE3) for 20 h with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).
335 Cells were resuspended in buffer A (50 mM Tris pH7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Guanidinium
336 HCl, protease inhibitors and 10 mM 2-Mercaptoethanol) and frozen in liquid N₂. Cells were thawed and
337 disrupted using an Avestin cell disruptor and centrifuged at 20K, the supernatant was treated with
338 protamine sulfate for 1 h at 4 °C and centrifuged again. The supernatant was loaded onto NiNTA beads
339 and incubated for 2 h at 4 °C. Beads were later washed with 10 column volumes (CVs) of buffer A, 10 CVs
340 buffer A containing 1M NaCl, 5 CVs buffer A and 5 CVs buffer B (50 mM Tris pH 7.5, 250 mM NaCl, 10 mM
341 Imidazole, 6 M Urea and 10 mM 2-Mercaptoethanol). Proteins were eluted with buffer B containing 150
342 mM Imidazole, and subjected to anion exchange on a HiTrap Q column in 20mM Tris pH7.4 and 6M Urea.
343 Purified proteins were concentrated and diluted 10x in buffer C (20 mM Tris pH7.5, 200 mM NaCl, 1 mM
344 TCEP and 0.1 mM PMSF), concentrated again and dialyzed overnight at 4 °C in buffer C.

345 His₆-AR-DBD-LBD (aa 554-919) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for
346 20 h with 0.2 mM isopropyl IPTG, with 10 μM of DHT and 10 μM ZnCl₂ added to the culture media. Cells
347 were re-suspended in buffer A (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 10% glycerol, 10 μM zinc acetate, and
348 10 μM DHT) and lysed using an Avestin cell disruptor, centrifuged at 20K and the supernatant incubated
349 with Ni-NTA beads for 2h at 4 °C. The resin was washed with 500 ml of wash buffer (20 mM Tris-HCl pH
350 7.5, 0.2 M NaCl, 10% glycerol, 10 μM zinc acetate, 10 μM DHT, and 20 mM imidazole), treated with 20U/ml
351 of Benzonase for 1 h at RT and then eluted with wash buffer containing 250 mM imidazole. TEV protease
352 was added to the eluted protein and dialyzed overnight against 20 mM Tris-HCl pH 7.5, 150 mM NaCl,

353 10% glycerol, 10 μ M ZnCl₂ and 10 μ M DHT to remove the His6-tag. The protein was further purified using
354 a Hiloal 16/60 Superdex S200 gel filtration column in 20 mM Tris-HCl, pH 7.5, 150mM NaCl, 5% glycerol,
355 1 mM TCEP and 10 μ M DHT.

356 Glutathione-S-transferase (GST)-AR-LBD (aa 663-919) was expressed with a pGEX-KG vector at 15
357 °C in *E. coli* BL21 (DE3) for 18-24 h with 0.06 mM IPTG in LB media with 10 μ M DHT. Cells were re-
358 suspended in 50 mM Tris pH 7.2, 150 mM NaCl, 2 mM DTT, 10% Glycerol, protease inhibitor cocktail, 5
359 mM EDTA and 10 μ M DHT. Cells were lysed on an Avestin cell disruptor and spun at 20K and incubated
360 on a glutathione agarose-resin at 4 °C overnight. The resin was washed with 5 CVs washing buffer (50 mM
361 Tris, pH 8.0, 150 mM NaCl, 5 mM EDTA, 10% Glycerol, 10 μ M DHT, 1 mM DTT), washed with 5 CVs
362 benzonase buffer and incubated with 20U/ml of Benzonase at RT for 1-2 h. The resin was washed with
363 TCB (50mM Tris, pH 8.0, 150 mM NaCl, 2.5 mM CaCl₂, 10 μ M DHT, 10% Glycerol, 0.1% beta-Octyl
364 Glucoside). Eluted protein was further purified by gel filtration on a Hiloal16/60 Superdex S200 column
365 in 25 mM Tris pH 7.4, 125 mM NaCl, 1 mM TCEP, 10 μ M DHT, 5% glycerol, 0.1% beta-Octyl glucoside.

366 His₆-AR-NTD-DBD (aa 1-632) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for
367 20 h with 0.2 mM IPTG. Cells were re-suspended in buffer A (50 mM Tris pH7.5, 250 mM NaCl, 10 mM
368 Imidazole, protease inhibitors and 10 mM 2-Mercaptoethanol) and frozen in liquid N₂. Cells were thawed
369 and disrupted using an Avestin cell disruptor and centrifuged at 20K and supernatant incubated with Ni-
370 NTA beads for 2 h at 4 °C. Beads were washed with 20 CVs of re-suspension buffer and treated with
371 20U/ml of Benzonase for 1 h at RT with rotation. Protein was eluted with re-suspension buffer containing
372 250 mM imidazole and subjected to Gel-filtration chromatography on a Superdex S200 column in 50 mM
373 Tris, 150 mM NaCl and 1 mM TCEP.

374 GST-KDM4A-CD (aa 1-350), GST-KG-KDM4B-CD (aa 1-348) and GST-KG-KDM4C-CD (aa 1-350)
375 were expressed with pGEX-KG vectors at 18 °C in *E. coli* BL21 (DE3) for 20 h with 0.2 mM IPTG in LB media
376 with 0.1 mM ZnCl₂ and 0.1 mM FeSO₄. Cells were re-suspended in 50 mM Tris pH 7.2, 300 mM NaCl, 2 mM

377 DTT and protease inhibitors. After fast freezing in liquid N₂, cells were broken on an Avestin cell disruptor,
378 spun at 20K and loaded on glutathione agarose-resin at 4°C overnight. The resin was extensively washed
379 with 10 CVs phosphate buffer saline (PBS), 10 CVs PBS containing 1 M NaCl, 5 CVs PBS and 5 CVs Benzonase
380 buffer (25 mM Tris pH 8.0, 50 mM NaCl, 2 mM MgCl₂), and treated with 20 U/ml of Benzonase for 1 h at
381 RT with slow rotation. The resin was then washed with 10 CVs PBS containing 1M NaCl, 10 CVs PBS and 5
382 CVs of thrombin cleavage buffer, then incubated with thrombin to remove the GST-tag for 3 h at RT. The
383 protein was purified by Cation exchange chromatography on a HiTrap S column with 50 mM MES buffer
384 pH6.5, followed by size exclusion chromatography on a Hiloal 16/60 Superdex S75 column in 25 mM Tris
385 pH 7.4, 125 mM NaCl and 1 mM TCEP.

386 His₆-KDM4A(301-708) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for 20 h
387 with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were re-suspended in buffer A (50 mM
388 Tris pH 7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Guanidinium HCl, protease inhibitors and 10 mM 2-
389 Mercaptoethanol). Cells were disrupted and the supernatant was loaded onto NiNTA beads and washed
390 with 10 CVs buffer A, 10 CVs buffer A containing 1M NaCl, 5 CVs buffer A and 5 CVs buffer B (50 mM Tris
391 pH 7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Urea and 10 mM 2-Mercaptoethanol). Proteins were eluted
392 with buffer B containing 150 mM Imidazole and further purified by gel filtration on a S75 column in 20
393 mM Tris pH7.5, 200 mM NaCl, 1 mM TCEP.

394 His₆-KDM4A(703-1064) was expressed with a pET28 vector at 16 °C in *E. coli* BL21 (DE3) for 20 h
395 with 0.5 mM IPTG in LB media and pellet re-suspended in buffer RB (50mM Tris pH=7.5, 250 mM NaCl 4
396 mM Imidazole, protease inhibitors). After protein binding, the Ni-NTA beads were washed with RB and
397 treated with Benzonase in 10 ml Benzonase buffer for 2h RT, washed again with RB and eluted with elution
398 buffer (50 mM Tris pH 7.5, 250 mM NaCl 250-500 mM Imidazole). Proteins were then subjected to
399 exclusion chromatography on a S200 column in 25 mM Tris pH 7.0, 125 mM NaCl and 1 mM TCEP.

400 GST-AR (93-495) was expressed with a pGEX-KG vector at 16 °C in *E. coli* BL21 (DE3) for 20 h with
401 0.5 mM IPTG in LB media. Cells were re-suspended in PBS, protease inhibitors cocktail and 5 mM DTT.
402 Cells were broken on an Avestin cell disruptor, spun at 20K and incubated with glutathione agarose-resin
403 at 4°C overnight. The resin was extensively washed with 10 CVs PBS, 10 CVs PBS containing 1 M NaCl, 5
404 CVs PBS and 5 CVs Benzonase buffer (25mM Tris pH8.0, 50mM NaCl, 2 mM MgCl₂), and treated with
405 20U/ml of Benzonase for 1-2 h at RT with rotation. The beads were washed with 5 CVs PBS and treated
406 with 20U/ml of TEV protease to remove the GST tag overnight at 4°C. Collected fractions were subjected
407 to anion exchange chromatography on a HiTrap-Q column with 20mM Bis-Tris buffer pH 6.4 and size
408 exclusion chromatography on a S75 column in 25 mM Tris pH 7.4, 125 mM NaCl and 2 mM TCEP.

409 To generate uniformly ¹⁵N-labeled proteins for NMR studies, bacteria were grown in M9 minimal
410 medium supplemented with ¹⁵NH₄Cl (CIL, Andover, MA) as the sole nitrogen source. Proteins were purified
411 as described above.

412

413 **NMR spectroscopy**

414 All NMR spectra were acquired at 25 °C on a Varian INOVA 600 MHz spectrometer. Samples for ¹H-¹⁵N
415 TROSY-HSQC measurements contained 40-50 μM of the ¹⁵N labelled protein and unlabeled protein.
416 Experiments with ¹⁵N-KDM4A(1-350) or ¹⁵N-KDM4B(1-348) were performed in 25 mM Tris (pH 7.4), 125
417 mM NaCl and 1 mM TCEP with 5% D₂O. For samples containing ¹⁵N-AR-LBD or ¹⁵N-KDM4C(1-350), the
418 buffer was 25 mM Tris (pH 7.4), 125 mM NaCl, 0.1% Octyl β-D-glucopyranoside and 1 mM TCEP with 5%
419 D₂O. For the NMR experiments including AR-LBD or AR-DBD+LBD, 1 μM DHT was included in the buffer.
420 Total acquisition times were 2–10 h. NMR data were processed with NMRPipe(50) and analyzed with
421 NMRView(51).

422

423 **Chemical Cross-linking**

424 Cross-linking reactions were performed with mixtures of 15 μ M AR NTD-DBD (aa 1-628) with 15 or 30 μ M
425 KDM4A(1-350), or with mixtures of 15 μ M KDM4A(1-350) with 15 μ M, 30 μ M or 60 μ M AR-NTD (aa 93-
426 495). The mixtures were incubated with 1 mM bis-(sulfosuccinimidyl)suberate (BS3) for 1 h at RT and the
427 reactions quenched with 25 mM Tris (pH 8.0). Samples of 18 μ g of protein were loaded onto SDS gels for
428 PAGE and coomassie blue staining.

429

430 **Gel formation and turbidity assays**

431 AR-NTD proteins containing 12Q, 20Q, 31Q or 49Qs at \sim 50 μ M concentration were dialyzed against a
432 gelation buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA and 0.1 mM
433 PMSF overnight. The proteins were then concentrated to 500 μ M and incubated at RT for 24-48 h until
434 gel formation was observed. For turbidity assays, protein samples in 25 mM Tris pH 7.5, 150 mM NaCl and
435 1 mM TCEP were concentrated to 100 μ M and then diluted in the same buffer to the various
436 concentrations shown in Figs. 3C,D. The proteins were incubated at 4 $^{\circ}$ C or RT for 1 h, and the OD at 400
437 nm was then measured.

438

439 **Luciferase Assay**

440 LNCaP and PC3 cells were co-transfected with an ARE-luc construct containing three androgen response
441 elements (ARE) ligated in tandem to a luciferase reporter(52), along with a control vector pcDNA3.1-FLAG
442 or an AR-expression vector (pcDNA3.1-FLAG-AR) containing different polyQ sequence in the presence or
443 absence of DHT. LNCaP cells were maintained in RPMI phenol red free with 5% charcoal stripped FBS
444 during the assay. Transfections were carried out using FuGENE HD (Promega, Catalog. E2311) for LNCaP
445 cells and Lipofectamine 3000 Kit for PC3 cells (Thermo Fisher, Catalog. L30000) following the instructions
446 from the manufacturer. Cells were treated with vehicle or 10 nM Dihydrotestosterone (DHT) 24 h after
447 transfection. Cell extracts were prepared 48 h after transfection and assayed for luciferase activity using

448 the Promega luciferase detection kit. Luciferase activities were normalized to co-transfected β -
449 galactosidase activity.

450

451 **Western blot analysis**

452 Protein samples from LNCaP cells used to measure luciferase activity (described in Fig. 4A) were loaded
453 into 4% to 15% SDS-PAGE (Bio-Rad) and subjected to electrophoretic analysis and subsequent blotting.
454 Nitrocellulose membranes were incubated with the primary antibody (overnight at 4°C) and the relevant
455 secondary antibodies (1 h at room temperature). The antibodies were purchased from MilliporeSigma:
456 anti-AR (catalog 06-680); anti-FLAG M2 (catalog F3165) and anti- β -actin (catalog no. A5441). For the
457 experiments shown in Fig. 5C, LNCaP cells transiently transfected with Cherry-AR containing different
458 polyQ lengths were blotted against the same anti-AR and anti- β -actin antibodies.

459

460 **AR localization assessment**

461 LNCaP cells transiently transfected with Cherry-AR (full-length) containing different lengths of glutamine
462 repeat (12Q, 20Q, 31Q or 49Q) were seeded onto coverslips in RPMI-1640 with 10% Charcoal-stripped
463 fetal bovine serum (FBS), and continued to grow for 48 h under the same starvation conditions. Coverslips
464 were fixed with methanol for 10 min at -20°C, washed with PBS and mounted on slides with mounting
465 media containing 4',6-diamidino-2-phenylindole (DAPI). Cells were imaged as z-stacks using a widefield
466 Delta Vision Fluorescence microscope. Maximum projections were done before nuclear/cytoplasmic
467 localization of cherry-AR-FL was quantified using the ImageJ Macro "Intensity Ratio Nuclei Cytoplasm
468 Tool". The data were plotted as the percentage of nuclear and cytoplasmic distribution of Cherry AR. More
469 than 50 cells were analyzed per experimental condition.

470

471 **Puncta formation assay**

472 LNCaP cells stably expressing Cherry-AR containing different lengths of glutamine repeat (12Q, 20Q, 31Q
473 or 49Q) were seeded onto coverslips in RPMI-1640 with 10% Charcoal stripped FBS and continued to grow
474 for 48 h under the same starvation conditions. Twenty four h before fixation, the cells were treated with
475 10 nM DHT to induce AR nuclear localization. Coverslips were fixed with methanol for 10 min at -20 °C
476 and mounted on coverslips with mounting solution containing 4',6-diamidino-2-phenylindole (DAPI). A
477 widefield Delta Vision Fluorescence microscope was used to take z-stacks images through whole cells,
478 maximum projections were performed, and the puncta intensity of cherry-AR-FL was quantified using
479 ImageJ. More than 50 cells were analyzed per experiment per condition, and there were ~5-45 puncta per
480 cell. Intensity ratio (IR) was calculated as the ratio of fluorescent intensity in the puncta minus the
481 background divided by the fluorescent intensity in the nucleoplasm minus the background. The
482 background intensity was obtained from an image of an area where there were no cells.

483

484 **Data availability**

485 The datasets generated during and/or analyzed during the current study are available from the
486 corresponding author on reasonable request.

487

488 **Acknowledgments**

489 We thank Yilun Sun for technical assistance. This work was supported by grants from the Cancer
490 Prevention and Research Institute of Texas (CPRIT) (RP120717-AC and RP120717-P1 to Z.P.L.; RP120717-
491 P3 to J.R.), NIH (RO1 CA215063 to Z.P.L.), and from the Welch Foundation (I-1304 to J.R.).

492

493 References

- 494 1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin.* 2010;60(5):277-300.
- 495 2. Maximum androgen blockade in advanced prostate cancer: an overview of 22 randomised trials
496 with 3283 deaths in 5710 patients. Prostate Cancer Trialists' Collaborative Group. *Lancet.*
497 1995;346(8970):265-9.
- 498 3. Attard G, Cooper CS, de Bono JS. Steroid hormone receptors in prostate cancer: a hard habit to
499 break? *Cancer Cell.* 2009;16(6):458-62.
- 500 4. Eisenberger MA, Walsh PC. Early androgen deprivation for prostate cancer? *N Engl J Med.*
501 1999;341(24):1837-8.
- 502 5. Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R, et al. Molecular determinants of
503 resistance to antiandrogen therapy. *Nat Med.* 2004;10(1):33-9.
- 504 6. Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in
505 understanding mechanisms of resistance and optimizing androgen depletion. *Nat Clin Pract Urol.*
506 2009;6(2):76-85.
- 507 7. Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, et al. Nuclear receptor repression
508 mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell.* 1997;89(3):373-80.
- 509 8. Shang Y, Myers M, Brown M. Formation of the androgen receptor transcription complex. *Mol Cell.*
510 2002;9(3):601-10.
- 511 9. Yamane K, Toumazou C, Tsukada Y, Erdjument-Bromage H, Tempst P, Wong J, et al. JHDM2A, a
512 JmJc-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell.*
513 2006;125(3):483-95.
- 514 10. Wissmann M, Yin N, Muller JM, Greschik H, Fodor BD, Jenuwein T, et al. Cooperative
515 demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat Cell*
516 *Biol.* 2007;9(3):347-53.
- 517 11. Shin S, Janknecht R. Activation of androgen receptor by histone demethylases JMJD2A and
518 JMJD2D. *Biochem Biophys Res Commun.* 2007;359(3):742-6.
- 519 12. Perry AS, Watson RW, Lawler M, Hollywood D. The epigenome as a therapeutic target in prostate
520 cancer. *Nat Rev Urol.* 2010;7(12):668-80.
- 521 13. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, et al. Global histone modification patterns
522 predict risk of prostate cancer recurrence. *Nature.* 2005;435(7046):1262-6.
- 523 14. Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, et al.
524 Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res.*
525 2004;64(6):1975-86.
- 526 15. Simental JA, Sar M, Lane MV, French FS, Wilson EM. Transcriptional activation and nuclear
527 targeting signals of the human androgen receptor. *J Biol Chem.* 1991;266(1):510-8.
- 528 16. Doesburg P, Kuil CW, Berrevoets CA, Steketee K, Faber PW, Mulder E, et al. Functional in vivo
529 interaction between the amino-terminal, transactivation domain and the ligand binding domain of the
530 androgen receptor. *Biochemistry.* 1997;36(5):1052-64.
- 531 17. Wilson EM. Analysis of interdomain interactions of the androgen receptor. *Methods Mol Biol.*
532 2011;776:113-29.
- 533 18. Gelmann EP. Molecular biology of the androgen receptor. *J Clin Oncol.* 2002;20(13):3001-15.
- 534 19. Cao S, Zhan Y, Dong Y. Emerging data on androgen receptor splice variants in prostate cancer.
535 *Endocr Relat Cancer.* 2016;23(12):T199-T210.
- 536 20. Shao C, Yu B, Liu Y. Androgen receptor splicing variant 7: Beyond being a constitutively active
537 variant. *Life Sci.* 2019;234:116768.

- 538 21. Duan L, Chen Z, Lu J, Liang Y, Wang M, Roggero CM, et al. Histone lysine demethylase KDM4B
539 regulates the alternative splicing of the androgen receptor in response to androgen deprivation. *Nucleic*
540 *Acids Res.* 2019;47(22):11623-36.
- 541 22. Tan MH, Li J, Xu HE, Melcher K, Yong EL. Androgen receptor: structure, role in prostate cancer and
542 drug discovery. *Acta Pharmacol Sin.* 2015;36(1):3-23.
- 543 23. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene
544 mutations in X-linked spinal and bulbar muscular atrophy. *Nature.* 1991;352(6330):77-9.
- 545 24. Jochum T, Ritz ME, Schuster C, Funderburk SF, Jehle K, Schmitz K, et al. Toxic and non-toxic
546 aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures.
547 *Biochim Biophys Acta.* 2012;1822(6):1070-8.
- 548 25. Stanford JL, Just JJ, Gibbs M, Wicklund KG, Neal CL, Blumenstein BA, et al. Polymorphic repeats in
549 the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res.* 1997;57(6):1194-8.
- 550 26. Kumar R, Atamna H, Zakharov MN, Bhasin S, Khan SH, Jasuja R. Role of the androgen receptor
551 CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci.*
552 2011;88(13-14):565-71.
- 553 27. Chamberlain NL, Driver ED, Miesfeld RL. The length and location of CAG trinucleotide repeats in
554 the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.*
555 1994;22(15):3181-6.
- 556 28. Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, Yong EL. Long polyglutamine tracts in the androgen
557 receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *J*
558 *Clin Endocrinol Metab.* 1997;82(11):3777-82.
- 559 29. Irvine RA, Ma H, Yu MC, Ross RK, Stallcup MR, Coetzee GA. Inhibition of p160-mediated
560 coactivation with increasing androgen receptor polyglutamine length. *Hum Mol Genet.* 2000;9(2):267-74.
- 561 30. Lavery DN, McEwan IJ. Structural characterization of the native NH₂-terminal transactivation
562 domain of the human androgen receptor: a collapsed disordered conformation underlies structural
563 plasticity and protein-induced folding. *Biochemistry.* 2008;47(11):3360-9.
- 564 31. De Mol E, Fenwick RB, Phang CT, Buzon V, Szulc E, de la Fuente A, et al. EPI-001, A Compound
565 Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen
566 Receptor. *ACS Chem Biol.* 2016;11(9):2499-505.
- 567 32. De Mol E, Szulc E, Di Sanza C, Martinez-Cristobal P, Bertoncini CW, Fenwick RB, et al. Regulation
568 of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General
569 Transcription Regulators. *Structure.* 2018;26(1):145-52 e3.
- 570 33. Davies P, Watt K, Kelly SM, Clark C, Price NC, McEwan IJ. Consequences of poly-glutamine repeat
571 length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol*
572 *Endocrinol.* 2008;41(5):301-14.
- 573 34. Escobedo A, Topal B, Kunze MBA, Aranda J, Chiesa G, Mungianu D, et al. Side chain to main chain
574 hydrogen bonds stabilize a polyglutamine helix in a transcription factor. *Nat Commun.* 2019;10(1):2034.
- 575 35. Eftekharzadeh B, Piai A, Chiesa G, Mungianu D, Garcia J, Pierattelli R, et al. Sequence Context
576 Influences the Structure and Aggregation Behavior of a PolyQ Tract. *Biophys J.* 2016;110(11):2361-6.
- 577 36. Bouchard JJ, Otero JH, Scott DC, Szulc E, Martin EW, Sabri N, et al. Cancer Mutations of the Tumor
578 Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol Cell.*
579 2018;72(1):19-36 e8.
- 580 37. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low
581 complexity sequence domains form dynamic fibers within hydrogels. *Cell.* 2012;149(4):753-67.
- 582 38. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular
583 biochemistry. *Nat Rev Mol Cell Biol.* 2017;18(5):285-98.

- 584 39. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors
585 Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*.
586 2018;175(7):1842-55 e16.
- 587 40. Charbonnier S, Zanier K, Masson M, Trave G. Capturing protein-protein complexes at equilibrium:
588 the holdup comparative chromatographic retention assay. *Protein Expr Purif*. 2006;50(1):89-101.
- 589 41. Rizo J, Rosen MK, Gardner KH. Enlightening molecular mechanisms through study of protein
590 interactions. *J Mol Cell Biol*. 2012;4(5):270-83.
- 591 42. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of
592 aromatic residues determine the phase behavior of prion-like domains. *Science*. 2020;367(6478):694-9.
- 593 43. Peskett TR, Rau F, O'Driscoll J, Patani R, Lowe AR, Saibil HR. A Liquid to Solid Phase Transition
594 Underlying Pathological Huntingtin Exon1 Aggregation. *Mol Cell*. 2018;70(4):588-601 e6.
- 595 44. Seibel NM, Eljouni J, Nalaskowski MM, Hampe W. Nuclear localization of enhanced green
596 fluorescent protein homomultimers. *Anal Biochem*. 2007;368(1):95-9.
- 597 45. McEwan IJ. Intrinsic disorder in the androgen receptor: identification, characterisation and
598 drugability. *Mol Biosyst*. 2011.
- 599 46. Oppong E, Stier G, Gaal M, Seeger R, Stoeck M, Delsuc MA, et al. An Amyloidogenic Sequence at
600 the N-Terminus of the Androgen Receptor Impacts Polyglutamine Aggregation. *Biomolecules*. 2017;7(2).
- 601 47. Lyon AS, Peeples WB, Rosen MK. A framework for understanding the functions of biomolecular
602 condensates across scales. *Nat Rev Mol Cell Biol*. 2020.
- 603 48. Quayle SN, Mawji NR, Wang J, Sadar MD. Androgen receptor decoy molecules block the growth
604 of prostate cancer. *Proc Natl Acad Sci U S A*. 2007;104(4):1331-6.
- 605 49. Myung JK, Wang G, Chiu HH, Wang J, Mawji NR, Sadar MD. Inhibition of androgen receptor by
606 decoy molecules delays progression to castration-recurrent prostate cancer. *PLoS One*.
607 2017;12(3):e0174134.
- 608 50. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - A Multidimensional Spectral
609 Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr*. 1995;6(3):277-93.
- 610 51. Johnson BA, Blevins RA. Nmr View - A Computer-Program for the Visualization and Analysis of
611 Nmr Data. *Journal of Biomolecular Nmr*. 1994;4(5):603-14.
- 612 52. Roggero CM, Jin L, Cao S, Sonavane R, Kopplin NG, Ta HQ, et al. A detailed characterization of
613 stepwise activation of the androgen receptor variant 7 in prostate cancer cells. *Oncogene*.
614 2021;40(6):1106-17.

615

616

617

618

619 **Figure legends**

620

621 **Figure 1.** Low complexity sequences are involved in binding between AR and KDM4A. (A) Domain diagrams
622 of AR and KDM4A. Selected residue numbers indicating the protein N- or C-termini, or approximate
623 domain boundaries, are indicated above the diagrams. Domain names are shown below the diagrams (Tu
624 = Tudor). (B-E) Superpositions of ^1H - ^{15}N TROSY-HSQC spectra of ^{15}N -KDM4A(1-350) alone (black contours)
625 or in the presence of equimolar amounts of unlabeled AR-NTD (B), AR-NTD-DBD (C), AR-LBD (D) or AR-
626 DBD-LBD (E). (F-G) Superpositions of ^1H - ^{15}N TROSY-HSQC spectra of ^{15}N -AR-LBD alone (black contours) or
627 in the presence (red contours) of equimolar amounts of KDM4A(301-708) or KDM4A(703-1064).

628

629 **Figure 2.** The catalytic domain of KDM4A can be efficiently cross-linked with the AR N-terminal region. (A)
630 Samples containing 15 μM AR-NTD-DBD were incubated with 1 or 2 equivalents of KDM4A(1-350) and 1
631 mM BS3 for 1 h at RT, and the reactions were analyzed by SDS-PAGE followed by coomassie Blue staining
632 (two right-most lanes). The three lanes on the left show samples of AR-NTD-DBD and KDM4A(1-350) alone
633 or together without BS3. (B) Samples containing 15 μM KDM4A(1-350) were incubated with 1, 2 or 4
634 equivalents of AR(93-495) and 1 mM BS3 for 1 h at RT, and the reactions were analyzed by SDS-PAGE
635 followed by coomassie Blue staining (three right-most lanes). The three lanes on the left show samples of
636 KDM4A(1-350) and AR(93-495) alone or together without BS3. Note that the staining efficiency of AR
637 fragments containing the NTD with coomassie Blue is substantially lower than is usual in proteins because
638 of the paucity of basic residues in its sequence.

639

640 **Figure 3.** Longer polyQ sequences increase the tendency of AR-NTD to form gels. (A) Samples containing
641 AR-NTD with different polyQ length were concentrated to 600 μM (12Q), 520 μM (20Q), 520 μM (31Q)
642 and 480 μM (49Q) and incubated at RT for 24-48 h until gel formation was observed. Images of the

643 resulting samples are shown. (B) Images of a 500 μ M sample of AR-NTD-31Q that was allowed to form
644 gels for 48 h before (left) and after (right) adding 10% 1,6-hexanediol (HD). (C-D) Turbidity assays with
645 samples of AR-NTD containing different polyQ length. Samples were concentrated to 100 μ M,
646 immediately diluted to the indicated concentrations and incubated for 1 h at 4 $^{\circ}$ C (B) or RT (C). The OD at
647 400 nm was then measured. Bars show averages of values measured in three independent experiments
648 performed under the same conditions and error bars show standard deviations.

649

650 **Figure 4.** Longer polyQ sequences decrease the transcriptional activity of AR. (A,C) Luciferase reporter
651 assays were performed with LNCaP cells and PC3 cells transfected with vectors expressing full-length AR
652 containing 12, 20, 31 or 49 glutamines in the polyQ repeat. Luciferase activity was measured in the
653 absence and presence of DHT, and the activities were normalized to co-transfected β -galactosidase
654 activity. Bars show averages of values measured in three independent experiments performed under the
655 same conditions and error bars show standard deviations. (B) Western blot showing the levels of AR in
656 the LNCaP cells used for the assays shown in panel (A). Top: total AR using an anti-AR antibody; middle:
657 overexpressed FLAG-AR using an anti-Flag antibody; and bottom: anti- β -actin used as a loading control.

658

659 **Figure 5.** Longer polyQ sequences decrease the nuclear localization of AR in the absence of hormone. (A)
660 Representative images of LNCaP cells transiently transfected with Cherry-AR containing different lengths
661 of glutamine repeat (12Q, 20Q, 31Q or 49Q), in the absence of DHT. Red corresponds to Cherry
662 fluorescence and blue to DAPI fluorescence. (B) AR distribution in the nucleus (grey bars) or the cytoplasm
663 (red bars) normalized to the total in cells transiently transfected with Cherry-AR-12Q, -20Q, -31Q or -49Q.
664 More than 50 cells were analyzed for each condition. Bars show averages and error bars show standard
665 deviations. (C) Western blot showing the expression levels of Cherry-AR and endogenous AR proteins
666 using an anti-AR antibody and anti- β -actin antibody as a loading control.

667

668 **Figure 6.** Longer polyQ sequences increase the tendency of AR to localize to nuclear puncta in the
669 presence of DHT. (A) Representative images of cells stably expressing Cherry-AR-12Q, -20Q, -31Q or -49Q,
670 treated with DHT 24 h before fixation. Red corresponds to Cherry fluorescence and blue to DAPI
671 fluorescence. (B) Puncta intensity ratios (IRs) in cells stably expressing Cherry-AR-12Q, -20Q, -31Q or -49Q
672 were calculated as the ratio of fluorescent intensity in the puncta minus the background divided by
673 fluorescent intensity in the nucleoplasm minus the background. The background intensity corresponds to
674 an image taken from an area where there were no cells. Bars represent average IRs calculate from
675 measurements performed in at least 50 cells under each condition (5-45 puncta per cell) and errors bars
676 show standard deviations. Statistical significance and P values were determined by one-way analysis of
677 variance (ANOVA) with Holm-Sidak test (** P < 0.01; *** P < 0.001).

678

679

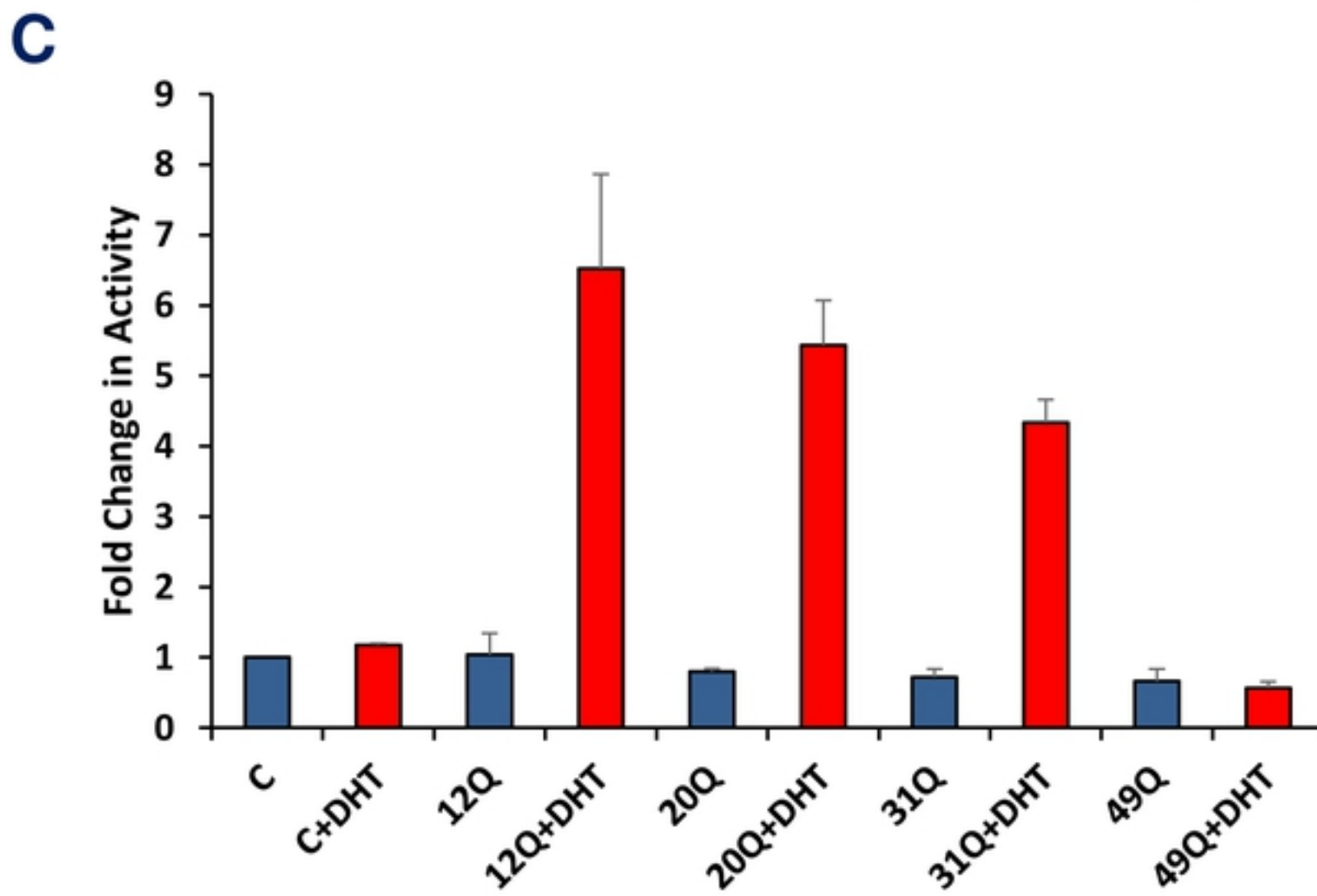
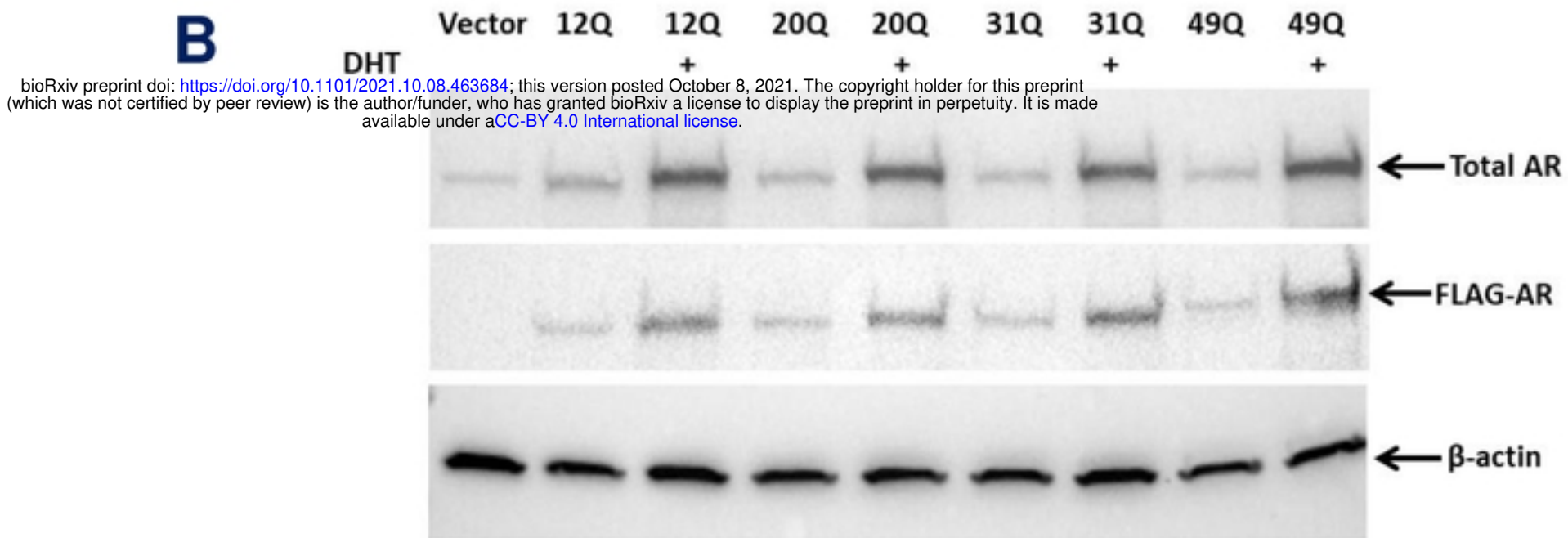
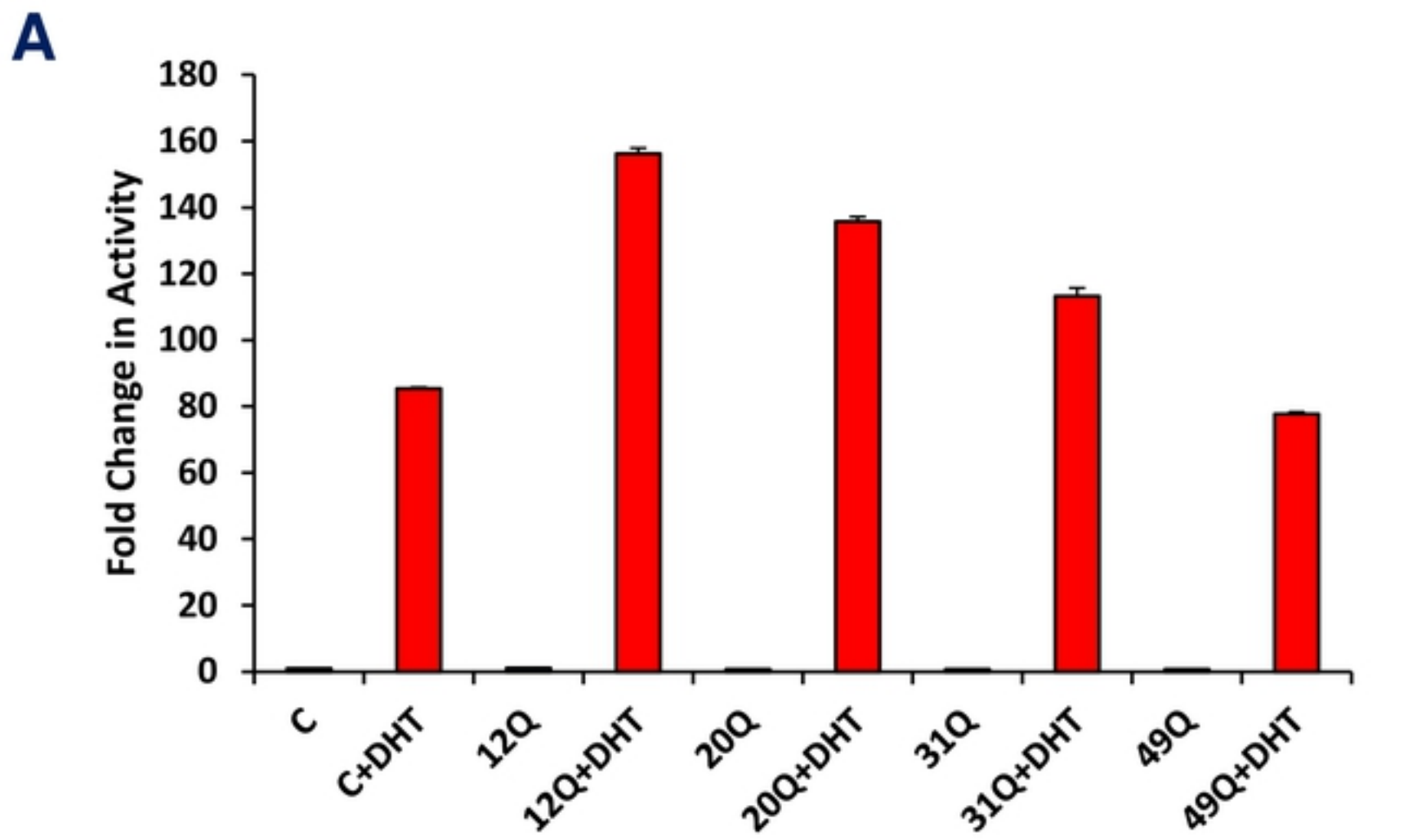


Figure 4
Roggero et al.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.10.08.463684>; this version posted October 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

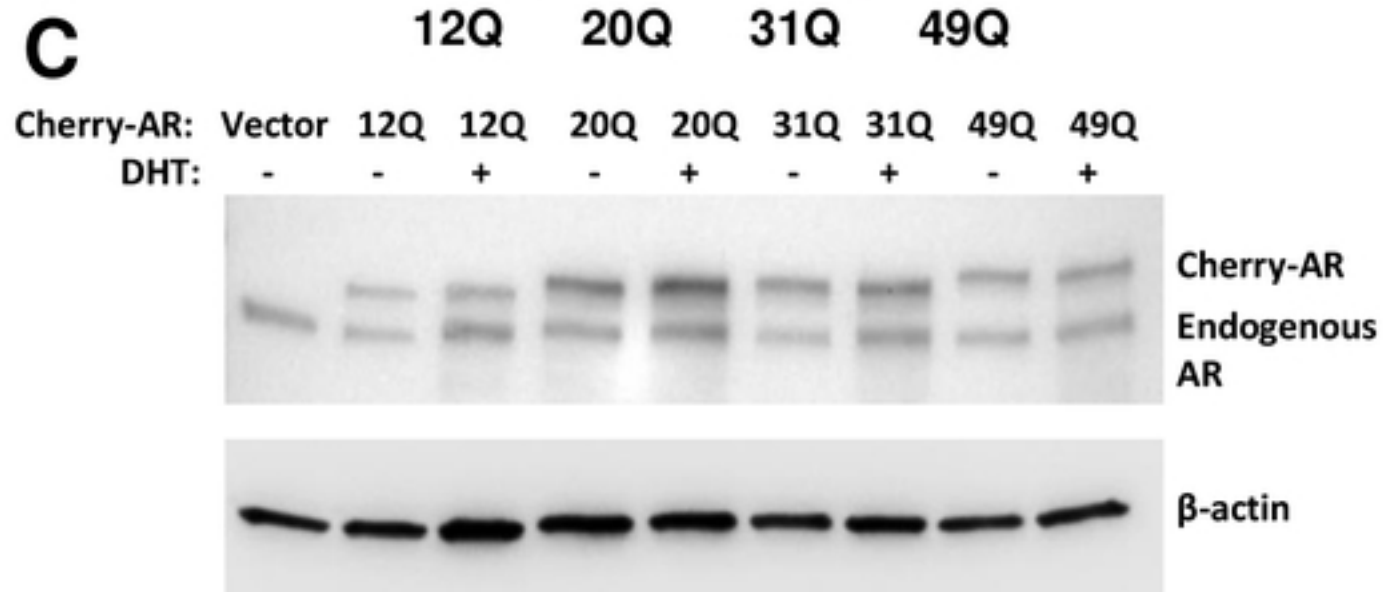
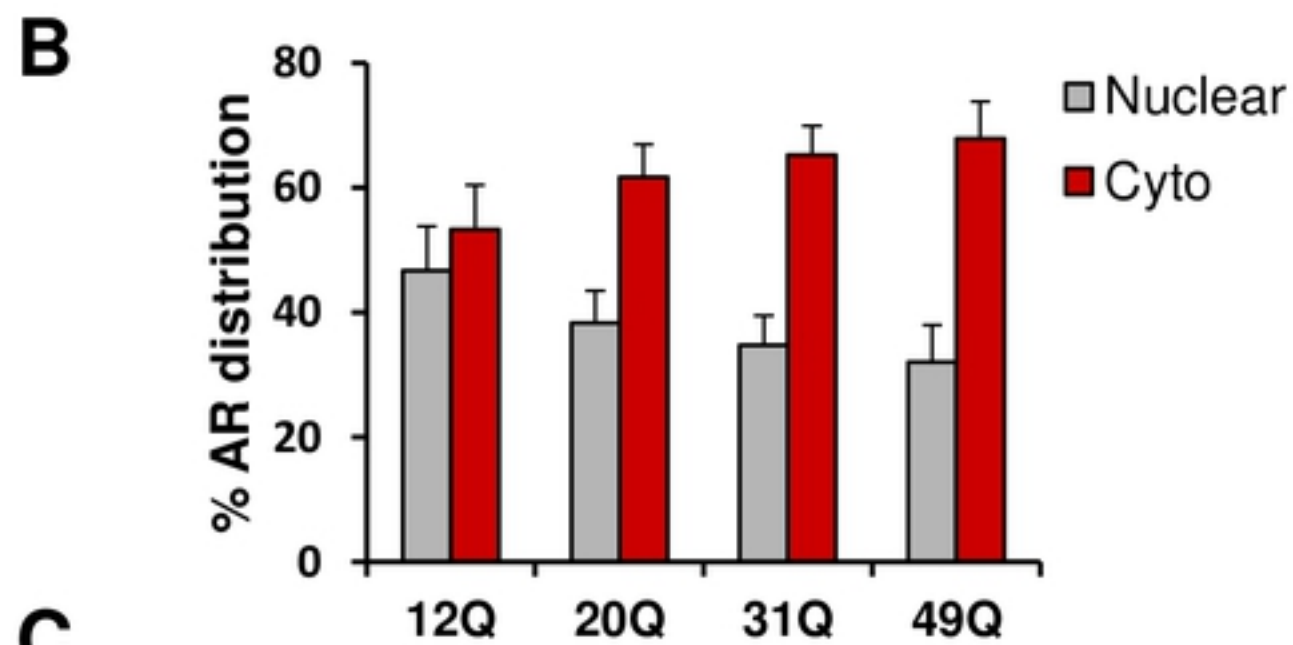
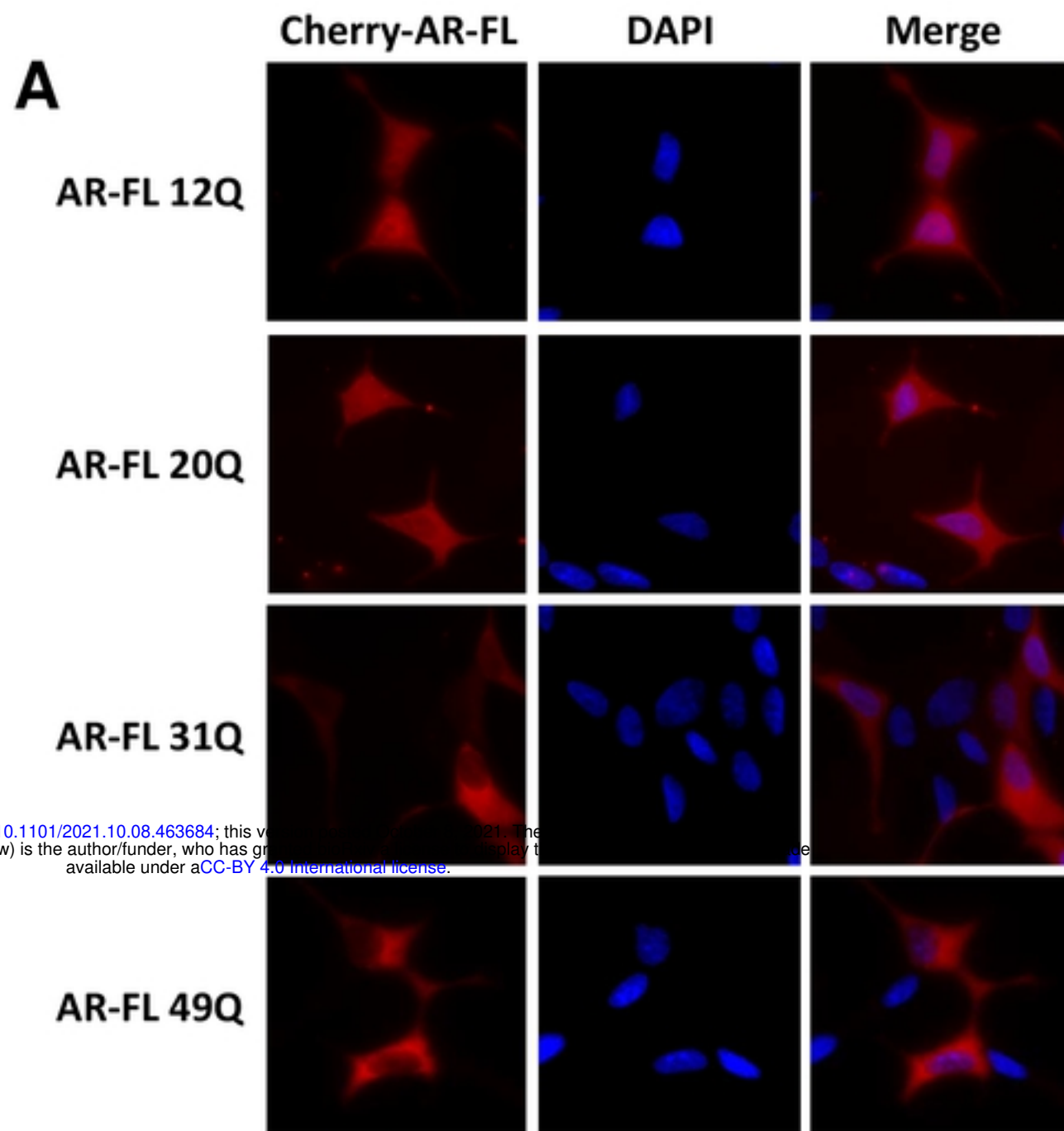


Figure 5
Roggero et al.

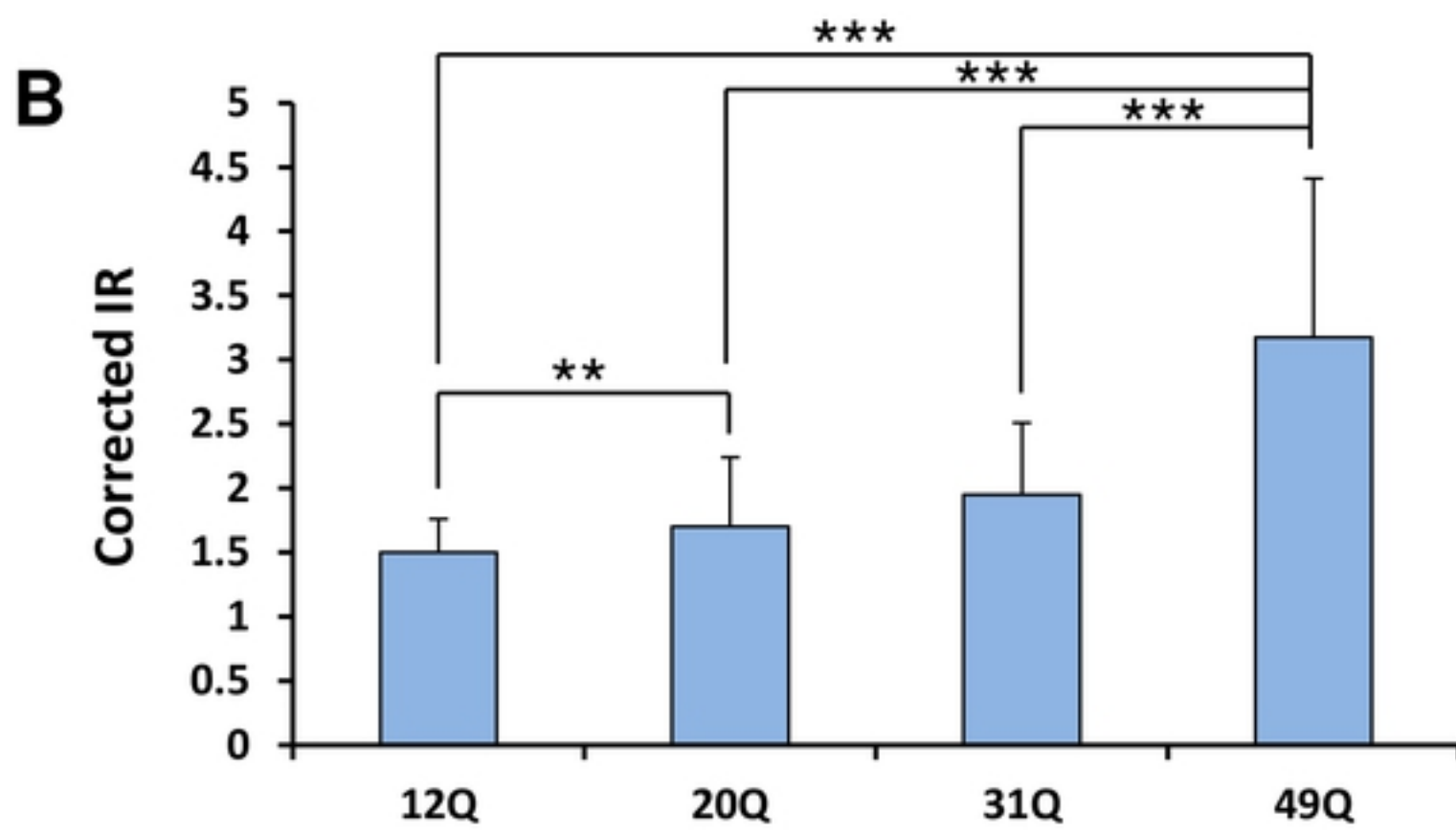
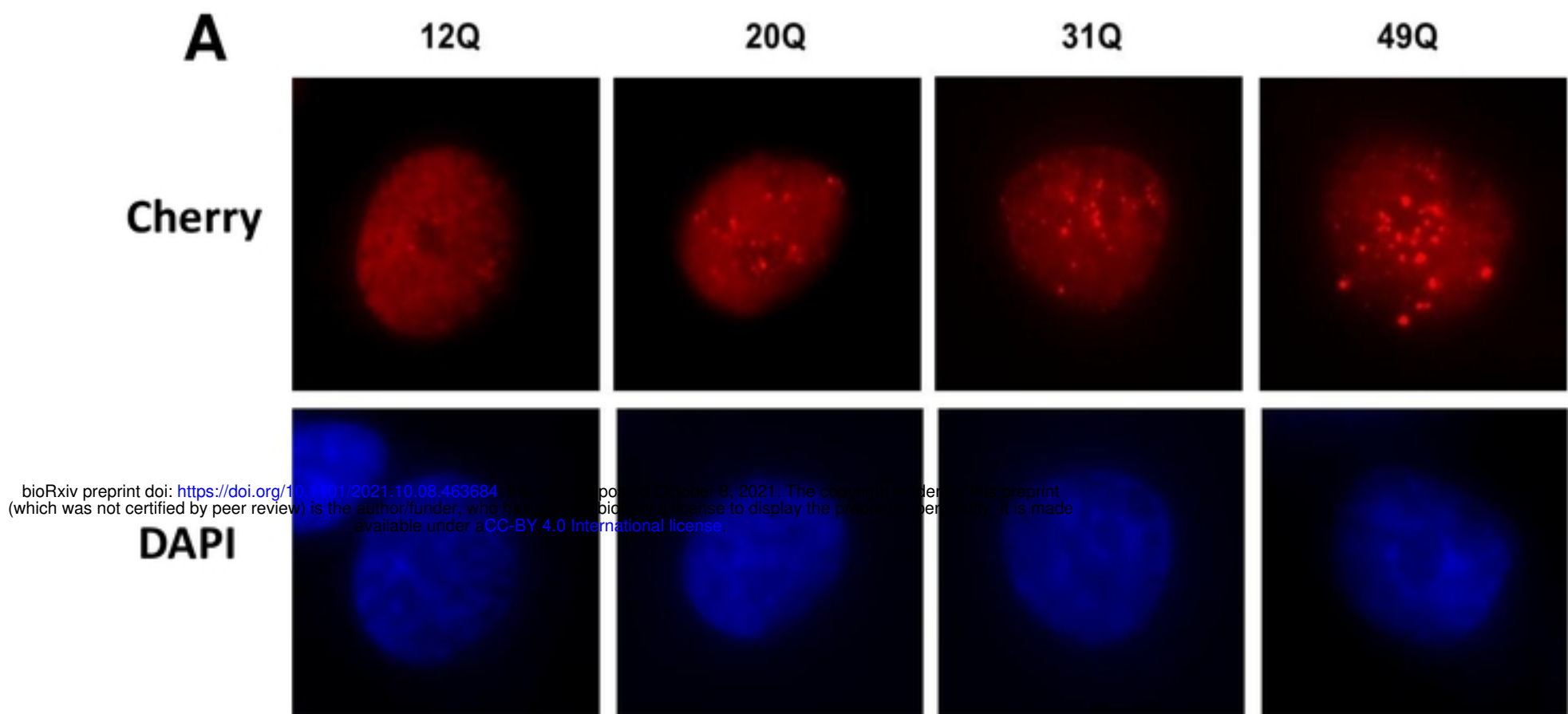


Figure 6
Roggero et al.

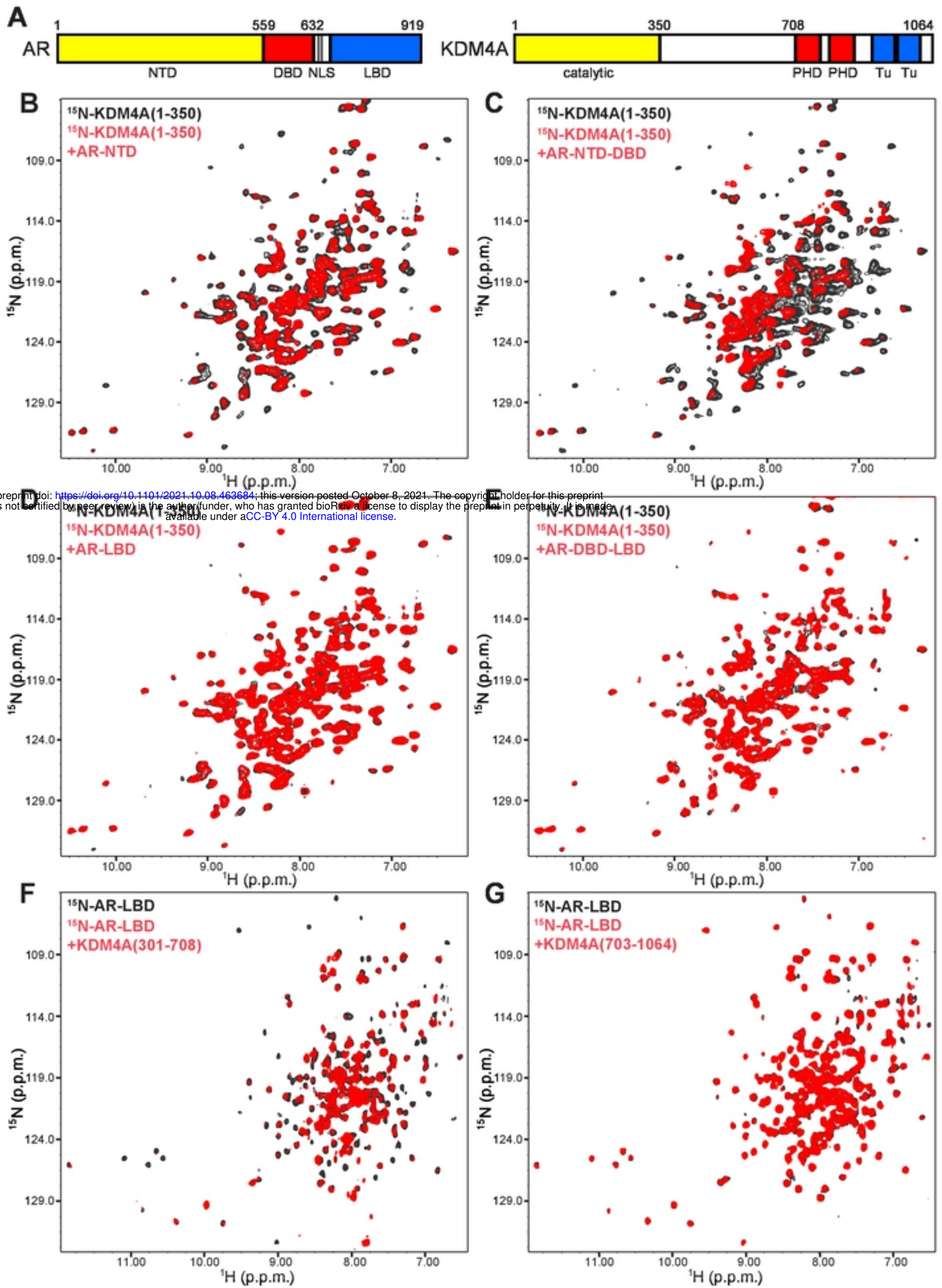


Figure 1
Roggero et al.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.10.08.463684>; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

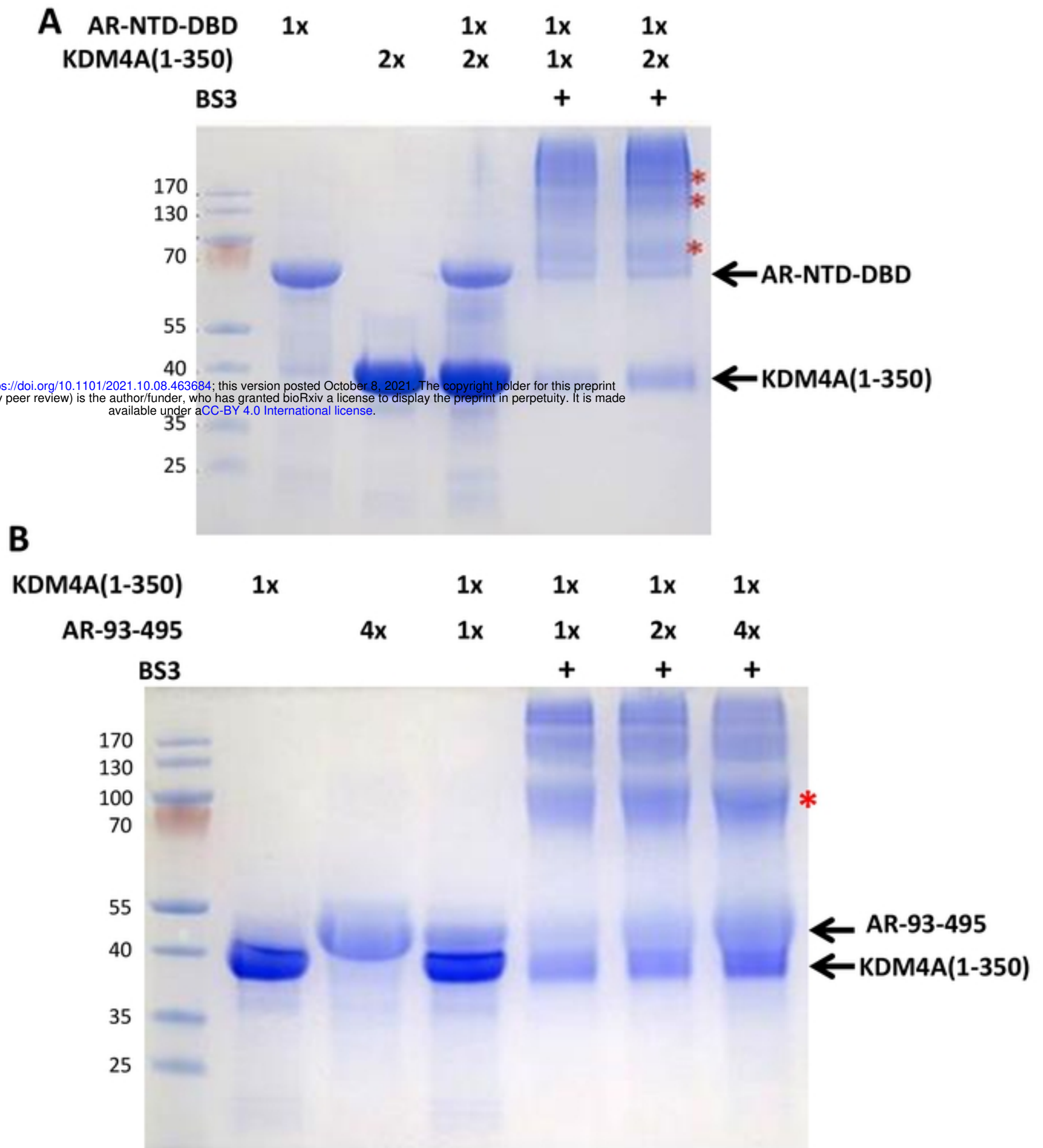


Figure 2
Roggero et al.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.10.08.463684>; this version posted October 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

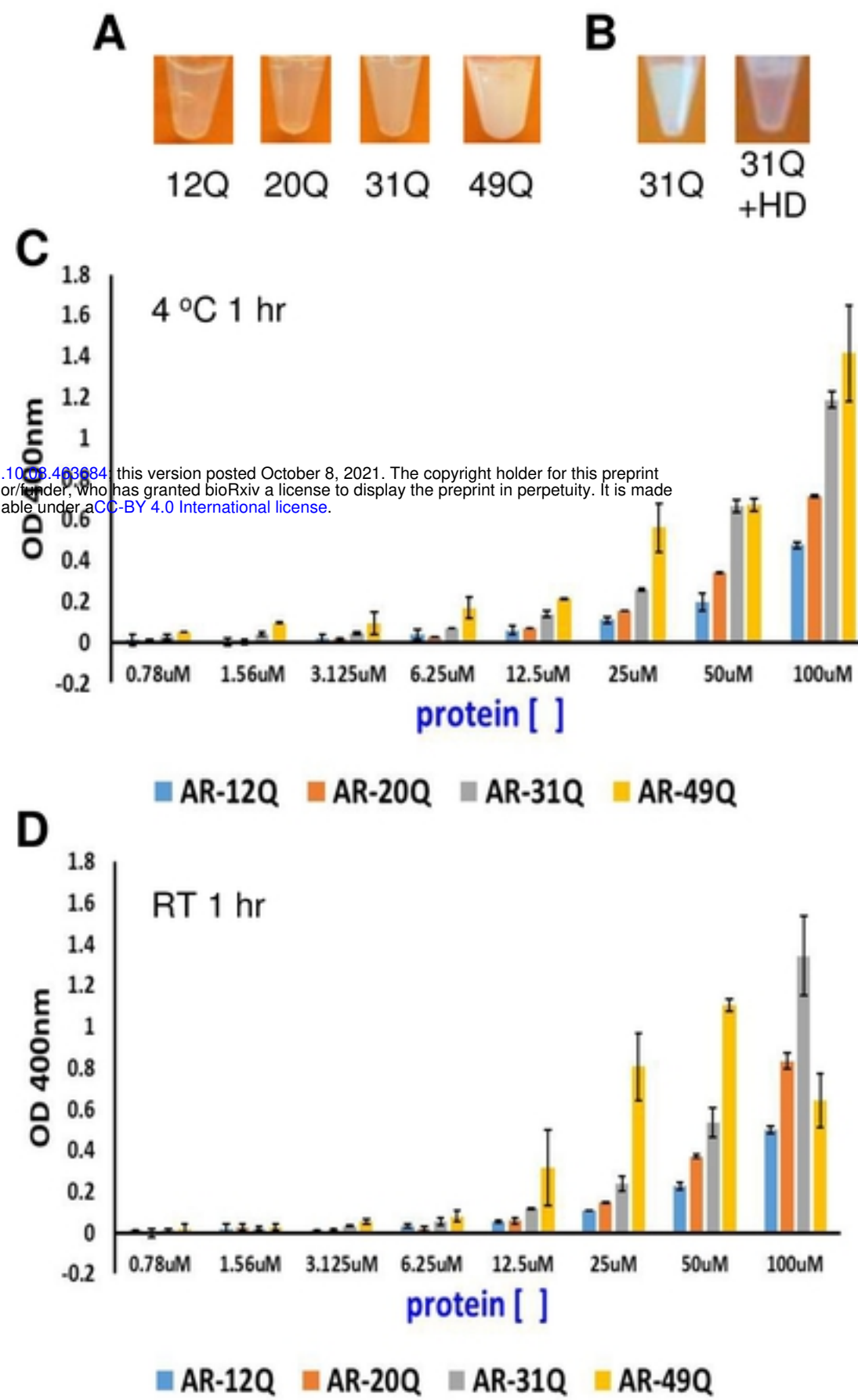


Figure 3
Roggero et al.