Phylogenomics-guided discovery of a novel conserved cassette of short linear motifs in BubR1 essential for the spindle checkpoint

Eelco Tromer[1,4], Debora Bade[1], Berend Snel[4 *], Geert J.P.L. Kops[1,2,3 *]

[1] *Hubrecht Institute – KNAW (Royal Netherlands Academy of Arts and Sciences) Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.*

[2] *Cancer Genomics Netherlands, and* [3] *Center for Molecular Medicine, University Medical Center Utrecht, 3584 CG, Utrecht, The Netherlands.*

[4]*Theoretical Biology and Bioinformatics, Department of Biology, Science Faculty, Utrecht University, 3584 CH Utrecht, The Netherlands.*

*Corresponding authors: g.kops@hubrecht.eu or b.snel@uu.nl

Abstract

The spindle assembly checkpoint (SAC) maintains genomic integrity by preventing progression of mitotic cell division until all chromosomes are stably attached to spindle microtubules. The SAC critically relies on the paralogs Bub1 and BubR1/Mad3, which integrate kinetochore-spindle attachment status with generation of the anaphase inhibitory complex MCC. We previously reported on the widespread occurrences of independent gene duplications of an ancestral 'MadBub' gene in eukaryotic evolution and the striking parallel subfunctionalization that lead to loss of kinase function in BubR1/Mad3-like paralogs. We now present an elaborate subfunctionalization analysis that includes all known motifs in Bub1 and BubR1, and show that ancestral features are consistently retained in the same functional paralog: GLEBS/CDI/CDII/kinase in the Bub1-like and KEN1/KEN2/D-Box in the BubR1/Mad3-like. The recently described ABBA motif can be found in either or both paralogs. We however discovered two additional ABBA motifs that flank KEN2. This cassette of ABBA1-KEN2-ABBA2 forms a strictly conserved module in all ancestral and BubR1/Mad3-like proteins, suggestive of a specific and crucial SAC function. Indeed, deletion of the ABBA motifs in human BUBR1 abrogates the SAC and affects APC/C-Cdc20 interactions. Our detailed comparative genomics analyses thus enabled discovery of a conserved cassette of motifs essential for the SAC and shows how this approach can be used to uncover hitherto unrecognized functional protein features.

Introduction

Chromosome segregation during cell divisions in animals and fungi is monitored by a cell cycle checkpoint known as the spindle assembly checkpoint (SAC) (Vleugel et al., 2012; London and Biggins, 2014a; Musacchio, 2015). The SAC couples absence of stable attachments between kinetochores and spindle microtubules to inhibition of anaphase by assembling a four-subunit inhibitor of the anaphase-promoting complex (APC/C), known as the MCC (Sacristan and Kops, 2015; Izawa and Pines, 2014; Chao et al., 2012). The molecular pathway that senses lack of attachment and produces the MCC relies on two related proteins known as Bub1 and BubR1/Mad3 (London and Biggins, 2014a). Bub1 is a serine/threonine kinase that localizes to kinetochores and promotes recruitment of MCC subunits and of factors that stimulate its assembly (Klebig et al., 2009; London and Biggins, 2014b; Vleugel et al., 2015) These events are largely independent of Bub1 kinase activity, however, which instead is essential for the correction process of attachment errors (Kawashima et al., 2010; Klebig et al., 2009; Andrews et al., 2004). BubR1/Mad3 is one of the MCC subunits, responsible for directly preventing APC/C activity and anaphase onset (Tang et al., 2001; Sudakin et al., 2001; Chao et al., 2012). It does so by contacting multiple molecules of the APC/C co-activator Cdc20, preventing APC/C substrate access and binding of the E2 enzyme UbcH10 (Chao et al., 2012; Izawa and Pines, 2014; Alfieri et al., 2016; Yamaguchi et al., 2016). The BubR1/Mad3-Cdc20 contacts occur via various short linear motifs (SLiMs) known as ABBA, KEN, and D-box (Alfieri et al., 2016; Chao et al., 2012; Lischetti et al., 2014; King et al., 2007; DiFiore et al., 2015; Burton and Solomon, 2007b). Like Bub1, BubR1 also impacts on the attachment error-correction process via a KARD motif that recruits the PP2A-B56 phosphatase (Suijkerbuijk et al., 2012b; Kruse et al., 2013; Xu et al., 2013).

This may not however be a universal feature of BubR1/Mad3-like proteins, because many lack a KARD-like motif.

Bub1 and BubR1/Mad3 are paralogs. We previously showed they originated by similar but independent gene duplications from an ancestral MadBub gene in many lineages, and that the two resulting gene copies then subfunctionalized in remarkably comparable ways (Suijkerbuijk et al., 2012a). An ancestral N-terminal KEN motif (KEN1: essential for the SAC) and an ancestral C-terminal kinase domain (essential for attachment error-correction) were retained in only one of the paralogous genes in a mutually exclusive manner in virtually all lineages (i.e. one gene retained KEN but lost kinase, while the other retained kinase but lost KEN). One exception to this 'rule' are vertebrates where both paralogs have a kinase-like domain. The kinase domain of human BUBR1 however lacks enzymatic activity (i.e. is a pseudokinase) but instead confers stability onto the BUBR1 protein (Suijkerbuijk et al., 2012a).

The similar subfunctionalization of Bub1 and BubR1/Mad3-like paralogs was inferred from analysis of two domains (TPR and kinase) and one motif (KEN1). We set out to analyze whether any additional features specifically segregated to Bub1- or BubR1/Mad3-like proteins after duplications by designing an unbiased feature discovery pipeline and tracing feature evolution. The pipeline extracted all known and various previously unrecognized conserved motifs from Bub1/BubR1 family gene members. Two of these are novel ABBA motifs that flank KEN2 specifically in BuBr1/Mad3-like proteins and we show that this highly conserved and ABBA-KEN2-ABBA cassette is crucial for the SAC in human cells.

Results and Discussion

Refined phylogenomic analysis of the MadBub gene family pinpoints 16 independent gene duplication events across the eukaryotic tree of life.

To enable detailed reconstruction of subfunctionalization events of all known functional features after duplication of ancestral MadBub genes, we expanded our previously published set of homologs (Suijkerbuijk et al., 2012a) through broader sampling of sequenced eukaryotic genomes, focusing on sequences closely associated with duplication events (**supplementary sequence file 1**). Phylogenetic analyses of a multiple sequence alignment of the TPR domain (the only domain shared by all MadBub family members) of 149 MadBub homologs (**supplementary discussion** and **supplementary figure 1**) corroborated the ten independent duplications previously described (Suijkerbuijk et al., 2012a) and allowed for a more precise determination of the age of the duplications. Strikingly, we found evidence for a number of additional independent duplications: Three duplications in stramenopile species of the SAR super group (*Albuginaceae* {#10 in **figure 1b**}, *E. siliculosis* {#11} and *A. anophagefferens* {#12}) and one at the base of basidiomycete fungi ({#4}, puccinioimycetes). The BUBR1 paralog in teleost fish underwent a duplication and fission event, of which the C-terminus product was retained only in the lineage leading to zebra fish (*D. rerio* {#7}). Lastly, through addition of recently sequenced genomes we could specify a duplication around the time plants started to colonize land ({#13}, bryophytes) and an independent duplication in the ancestor of higher plants ({#14}, tracheophytes), followed by a duplication in the ancestor of the flowering plants ({#15}, magnoliaphytes). These gave rise to three MadBub homologs, signifying additional subfunctionalization of the paralogs in the plant model organism *A. thaliana*. It thus seems to be the case that such

striking parallel subfunctionalization as we originally identified, is indeed predictive for more of its occurrence in lineages whose genome sequences have since been elucidated.

De novo discovery, phylogenetic distribution and fate after duplication of functional motifs in the MadBub gene family.

Previous analyses revealed a recurrent pattern of mutually exclusive retention of an N-terminal KEN-box and a C-terminal kinase domain after duplication of an ancestral MadBub (Suijkerbuijk et al., 2012a; Murray, 2012). These patterns suggested the hypothesis of paralog subfunctionalization towards either inhibition of the APC/C in the cytosol (retaining the KEN-box) or attachment-error correction at the kinetochore (retaining the kinase domain). Given the extensive sequence divergence of MadBub homologs and a scala of different known functional elements, we reasoned that a comprehensive analysis of MadBub gene duplicates would provide opportunities for the discovery of novel and co-evolving ancestral features. For clarity we refer to the Bub1-like paralog (C-terminal kinase domain) as BUB and the BubR1/Mad3-like paralog (N-terminal KEN box) as MAD throughout the rest of this manuscript.

To capture conserved ancestral features of diverse eukaryotic MadBub homologs, we constructed a sensitive de novo motif and domain discovery pipeline (ConFeaX: conserved feature extraction) similar to our previous approach used to characterize KNL1 evolution (Tromer et al., 2015). In short, the MEME algorithm (Bailey et al., 2009) was used to search for significantly similar gapless amino acid motifs, and extended motifs were aligned by MAFFT (Katoh and Standley, 2013). Alignments were modeled using HMMER (Eddy, 2011) and sensitive profile HMM searches were iterated and specifically optimized using permissive E-values/bit scores until convergence

(**methods** and **figure 1a**). Due to the degenerate nature of the detected short linear motifs, we manually scrutinized the results for incorrectly identified features and supplemented known motif instances, when applicable. We preferred ConFeaX over other de novo motif discovery methods (e.g. Davey et al., 2012 Nguyen Ba et al., 2012) as it does not rely on high quality full length alignment of protein sequences and allows detection of repeated or dynamic non-syntenic conserved features (which is a common feature for SLiMs). It is therefore better tuned to finding conserved features over long evolutionary distances in general and specifically in this case where recurrent duplication and subfunctionalization hamper conventional multiple sequence alignment based analysis.

ConFeaX identified known functional motifs and domains and in some cases extended their definition: KEN1 (Murray and Marks, 2001), KEN2 (Burton and Solomon, 2007a), GLEBS (Taylor et al., 1998), KARD (Suijkerbuijk et al., 2012b; Xu et al., 2013; Kruse et al., 2013), CDI (Klebig et al., 2009), D-box (Burton and Solomon, 2007a), CDII (a co-activator domain of BUB1 (Kang et al., 2008; Klebig et al., 2009)) and the recently discovered ABBA motif (termed ABBA3 in **figure 3**) (DiFiore et al., 2015; Lischetti et al., 2014) (**figure 1a**, **supplementary table II** and **supplementary sequence file 2**). The TPR and the kinase domain were annotated using profile searches of previously established models (Suijkerbuijk et al., 2012a) and excluded from de novo sequence searches. KEN1 and KEN2 could be discriminated by differentially conserved residues surrounding the core KEN box (**figure 1a**). Those surrounding KEN1 are involved in the formation of the helix-turn-helix motif that positions yeast Mad3 towards Cdc20 (Chao et al., 2012), while two pseudo-symmetrically conserved tryptophan residues with unknown function specifically defined KEN2. Furthermore, we found that the third

position of the canonical ABBA motif is often occupied by a proline residue and the first position in ascomycetes (fungi) is often substituted for a polar amino acid [KRN] (**figure 1a**), signifying potential lineage-specific changes in Cdc20-ABBA interactions. Last, we also discovered a novel motif predominantly associated with the MAD paralog in basidiomycetes, plants, amoeba and stramenopiles but not metazoa, which we termed conserved motif I (CMI) (**figure 1a**).

Projection of the conserved ancestral features onto the MadBub gene phylogeny provided a highly detailed overview of MadBub motif evolution (**figure 1b**, **supplementary figure 1b**). We found that the core functional motifs and domains (TPR, KEN1, KEN2, ABBA, D-box, GLEBS, CMI, CDI, CDII and kinase) are present throughout the eukaryotic tree of life, representing the core features that were likely part of the SAC signaling network in the last eukaryotic common ancestor (LECA). Of note are lineages (nematodes, flatworms (*S. mansoni*), dinoflagellates (*S. minutum*) and early-branching fungi (microsporidia and *C. coronatus*)) for which multiple features were either lost or considerably divergent (**supplementary figure 1b**). Especially interesting is *C. elegans* in both KEN boxes and the GLEBS domain appeared to have been degenerated (ceMAD=san-1) and the CDI domain is lost (ceBUB=bub-1), indicating extensive rewiring or a less essential role of the SAC in nematode species, as has been suggested recently (Davey and Morgan, 2016; Moyle et al., 2014).

Our motif discovery analyses revealed the CDC20/Cdh1-interacting ABBA motif to be much more abundant than the single instances that were previously reported for BUBR1 and BUB1 in humans (Lischetti et al., 2014; DiFiore et al., 2015). We observed three different contexts for the ABBA motifs (**figure 1b**, **supplementary figure 1b**): (1)

in repeat arrays (e.g. MAD of *P. patens*, basidiomycetes and stramenopiles), (2) in the vicinity of CDI (many instances) and/or D-box/KEN (e.g. human), and (3) as two highly conserved ABBA motifs flanking KEN2 (virtually all species). Because of the positional conservation of the latter, we have termed these, ABBA1 and ABBA2. Any additional ABBA motifs were pooled in the category 'ABBA-other'.

In order to track the fate of the features discovered using ConFeaX, we quantified their co-presences and absences, as a proxy for co-evolution, by calculating the Pearson correlation coefficient (r) for the profiles of each domain/motif pair of 16 duplicated MadBub homologs (**figure 1b**)(Wu et al., 2003). Subsequent average clustering of the Pearson distance ($d=1-r$) revealed two sets of co-segregating and anti-correlated conserved features (**figure 2a+b**) consistent with our hypothesis that MadBub gene duplication caused parallel subfunctionalization of features towards the kinetochore (mainly BUB) and the cytosol (MAD) (Suijkerbuijk et al., 2012a). GLEBS, CDI, ABBA-other, KARD, CDII and the kinase domain formed a coherent cluster of features with bona fide function at the kinetochore. For a detailed discussion on several intriguing observations regarding presence/absence of these motifs in several eukaryotic lineages, and what this may mean for BUB/MAD and SAC function in these lineages, see **Supplementary Discussion**. A second cluster contained known motifs that bind and interact with (multiple) CDC20 molecules, including KEN1, KEN2, and (to a lesser extent) the D-box. Our newly discovered ABBA motifs that flank KEN2 were tightly associated with KEN2 and KEN1 (**figure 2**). As such, the ABBA1-KEN2-ABBA2 cassette (**Figure 3a**) co-segregated with MAD function during subfunctionalization of MadBub gene duplicates. Although the D-box often co-occurs with the KEN-ABBA cluster, this motif was occasionally lost (*e.g.* archeaplastids, *S. pombe* and *A. anophagefferens*).

Finally, CMI co-segregated with the Cdc20-interacting motifs (Figure 2a), suggesting a MAD-specific role for this newly discovered motif (possibly in MCC function and/or Cdc20-binding) in species harboring it such as plants, basidiomycetes and stramenopiles.

The conserved ABBA1-KEN2-ABBA2 cassette is essential for proper SAC signaling in human cells

The strong correlation of the ABBA1-KEN2-ABBA2 cassette with KEN1 and the D-box, urged us to examine the role of these motifs in BUBR1-dependent SAC signaling in human cells. We therefore generated stable isogenic HeLa FlpIn cell lines expressing doxycyclin-inducible versions of LAP-tagged BUBR1 (Suijkerbuijk et al., 2010). These included: ΔABBA1, ΔABBA2, ΔABBA1+2, alanine-substitutions of the two KEN2-flanking tryptophans (W1-A, W2-A and W1/2-A), KEN1-AAA KEN2-AAA, ΔABBA3 and ΔD-box (**figure 3a-c**). The SAC was severely compromised in cells depleted of endogenous BUBR1 by RNAi, as measured by inability to maintain mitotic arrest upon treatment with S-trityl-L-cysteine (STLC) (Ogo et al., 2007) (median(*m*) = 50 minutes (min.) from NEBD to mitotic exit, compared control (*m*>500 min.)) (**figure 3d-e**). SAC proficiency was restored by expression of siRNA-resistant LAP-BUBR1 (*m*>500 min.). As shown previously (Burton and Solomon, 2007a; Elowe et al., 2010; Lara-Gonzalez et al., 2011), mutants of KEN1, KEN2 and the D-box strongly affected the SAC. Importantly, BUBR1 lacking ABBA1 or ABBA2 or both, or either of the two tryptophans, could not rescue the SAC (**figure 3e)**. We observed a consistently stronger phenotype for the mutated motifs on the N-terminal side of KEN2 (ΔABBA1 (*m*=65 min.) and W1-A (*m*=165 min.)) compared to those on the C-terminal side (ΔABBA2 (*m*=200 min.) and W2-A (*m*=260 min.)). The double ABBA (1/2) and tryptophan (1/2) mutants were however further

236 compromised (m = 50 and 110 min., respectively), suggesting non-redundant functions.
237 As expected from the interaction of ABBA motifs with the WD40 domain of CDC20 and
238 CDH1 (DiFiore et al., 2015), BUBR1 lacking ABBA1 and/or ABBA2 was less efficient in
239 binding APC/C-Cdc20 in mitotic human cells, to a similar extent as mutations in KEN1
240 (**figure 3f**). In our hands ABBA1 and 2 mutants were more strongly deficient in SAC
241 signaling and APC/C-Cdc20 binding than the previously described ABBA motif (ABBA3)
242 (**figure 3d-e)**. In conclusion therefore, the ABBA1-KEN2-ABBA2 cassette in BUBR1 is
243 essential for APC/C inhibition by the SAC.
244
245 We here discovered a symmetric cassette of SLiMs containing two Cdc20-binding ABBA
246 motifs and KEN2. This cassette strongly co-occurs with KEN1 in MAD-like and MadBub
247 proteins throughout eukaryotic evolution and has important contributions to the SAC in
248 human cells. Our co-precipitation experiments along with the known roles for ABBA-
249 like motifs and KEN2 and their recent modeling into the MCC-APC/C structure (Alfieri et
250 al., 2016; Yamaguchi et al., 2016) strongly suggest that the ABBA1-W-KEN2-W-ABBA2
251 cassette interacts with one or multiple Cdc20 molecules. Together with KEN1, these
252 interactions likely regulate affinity of MCC for APC/C or its positioning once bound to
253 APC/C. The constellation of interactions between Cdc20 molecules and the various
254 Cdc20-binding motifs in one molecule of BUBR1 (3x ABBA, 2x KEN and a D-box) is not
255 immediately obvious, and will have to await detailed atomic insights. The symmetric
256 arrangement of the cassette may be significant in this regard, as is the observation that
257 (despite a highly conserved WD40 structure of Cdc20) the length of spacing between
258 the ABBA motifs and KEN2 is highly variable between species. A more detailed
259 understanding of SAC function may be aided by ConFeaX-driven discovery of lineage-

specific conserved features in the MadBub family when more genome sequences become available, as well as of features in other SAC proteins families.

Supplementary information

Supplementary information included Supplementary discussion, 2 figures, 3 tables and 2 sequence files.

Methods

Phylogenomic analysis

We performed iterated sensitive homology searches with jackhammer (Finn et al., 2011) (based on the TPR, kinase, CDI, GLEBS and KEN boxes) using a permissive E-value and bitscore cut-off to include diverged homologs on uniprot release 2016_08 and Ensemble Genomes 32 (http://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer). Incompletely predicted genes were searched against whole genome shotgun contigs (wgs, http://www.ncbi.nlm.nih.gov/genbank/wgs) using tblastn. Significant hits were manually predicted using AUGUST (Stanke et al., 2006) and GENESCAN (Burge and Karlin, 1997). In total we used 152 MadBub homologs (**supplementary sequence file 1**). The TPR domains of 148 sequences were aligned using MAFFT-LINSI (Katoh and Standley, 2013); only columns with 80% occupancy were considered for further analysis. Phylogenetic analysis of the resulting multiple sequence alignment was performed using RAxML (Stamatakis, 2014) (**supplementary figure 1a**). Model selection was performed using Prot Test (Darriba et al., 2011) (Akaike Information Criterion): LG+G was chosen as the evolutionary model.

Conserved Feature Extraction and subfunctionalization analysis

ConFeaX starts with a probabilistic search for short conserved regions (max 50) using the MEME algorithm (option: any number of repeats) (Bailey et al., 2009). Significant motif hits are extended on both sides by 5 residues to compensate for the strict treatment of alignment information by the MEME algorithm. Next, MAFFT-LINSI (Katoh and Standley, 2013) introduces gaps and the alignments are modeled using the HMMER package (Eddy, 2011) and used to search for hits that are missed by the MEME

algorithm. Subsequent alignment and HMM searches were iterated until convergence. For short linear motifs with few conserved positions, specific optimization of the alignments and HMM models using permissive E-values/bit scores was needed (*e.g.* ABBA motif and D-box). Sequence logos were obtained using weblogo2 (Crooks et al., 2004). Subsequently, from each of the conserved features, a phylogenetic profile was derived (present is '1' and absent is '0') for all duplicated MadBub sequences as presented in figure 1. For all possible pairs, we determined the correlation using Pearson correlation coefficient (Wu et al., 2003). Average clustering based on Pearson distances (d=1-*r*) was used to indicate sub-functionalization.

Cell culture, transfection and plasmids

HeLa FlpIn T-rex cells were grown in DMEM high glucose supplemented with 10% Tet-free FBS (Clontech), penicillin/streptomycin (50 mg ml^{-1}), alanyl-glutamine (Sigma; 2 mM). pCDNA5-constructs were co-transfected with pOgg44 recombinase in a 10:1 ratio (Klebig et al., 2009) using FuGEHE HD (Roche) as a transfection reagent. After transfection, the medium was supplemented with puromycin (1 µg ml^{-1}) and blasticidin (8 µg ml^{-1}) until cells were fully confluent in a 10cm culture dish. siBUBR1 (5'-AGAUCCUGGCUAACUGUUCUU-3' custom Dharmacon) was transfected using Hiperfect (Qiagen) at 40 nM for 48h according to manufacturer's guidelines. RNAi-resistant LAP (YFP)-BUBR1 was sub cloned from plC58 (Suijkerbuijk et al., 2010) into pCDNA5.1-puro using AflII and BamHI restriction sites. To acquire mutants, site-directed mutagenesis was performed using the quickchange strategy (for primer sequences see **supplementary table III**).

Live cell imaging

316 For live cell imaging experiments, the stable HeLa-FlpIn-TRex cells were transfected
317 with 40nM siRNA (start and at 24 hrs). After 24 hrs, the medium was supplemented
318 with thymidine (2.5 mM) and doxycyclin (2 μg ml^{-1}) for 24 hrs to arrest cells in early S-
319 phase and to induce expression of the stably integrated construct, respectively. After 48
320 hrs, cells were released for 3 hrs and arrested in prometaphase of the mitotic cell cycle
321 (after ~8-10 hrs) by the addition of the Eg5 inhibitor S-trityl-L-cysteine (STLC, 20 μM).
322 HeLa cells were imaged (DIC) in a heated chamber (37 °C, 5% CO_2) using a CFI S Plan
323 Fluor ELWD 20x/NA 0.45 dry objective on a Nikon Ti-Eclipse wide field microscope
324 controlled by NIS software (Nikon). Images were acquired using an Andor Zyla 4.2
325 sCMOS camera and processed using NIS software (Nikon) and ImageJ.
326
327 *Immuno-precipitation and Western blot*
328 HeLa-FlpIn-TRex cells were induced with doxycyclin (2 μg ml^{-1}) 48 hrs before
329 harvesting. Synchronization by thymidine (2 mM) for 24 hrs and release for 10 hrs into
330 Taxol (2 μM) arrested cells in prometaphase. Cells were collected by mitotic shake-off.
331 Lysis was done in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5% NP40, 1mM EDTA, 1mM
332 DTT, protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktails 2 and 3
333 (Sigma). Complexes were purified using GFP-Trap beads (ChromoTek) for 15 min at
334 4°C. Precipitated proteins were washed with lysis buffer and eluted in 5x SDS sample
335 buffer. Primary antibodies were used at the following dilutions for western blotting:
336 BUBR1 (A300-386A Bethyl) 1:2000, alpha-Tubulin (T9026 Sigma) 1:5000, GFP
337 (Custom) 1:10 000, APC1 (A301-653A Bethyl) 1:2500, APC3 (gift from Phil Hieter)
338 1:2000, MAD2 (Custom) 1:2000, CDC20 (A301-180A Bethyl) 1:1000. Western blot
339 signals were detected by chemiluminescence using an ImageQuant LAS 4000 (GE
340 Healthcare) imager.

Author contributions

ET performed the motif search, phylogenetic analysis and SAC assays. DB performed the immunoprecipitation and western blot analyses. GK, BS conceived and managed the project. ET, BS and GK wrote the manuscript.

Acknowledgements

The authors thank the Snel and Kops lab for discussion and feedback. We thank Bas de Wolf and Laura Demmers for making cell lines. This work was supported by the UMC Utrecht and is part of the VICI research programme with project number 016.160.638, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). DB was supported by a DFG fellowship with number BA 5417/1-1, which is financed by the German Research Foundation (DFG).

Competing Financial Interests

The authors declare no competing financial interests.

Figure Legends

Figure 1 – Fate of conserved functional sequence features after 16 independent duplications of the MADBUB gene family throughout eukaryotic evolution (A) Overview of the de novo sequence discovery pipeline ConFeaX including the ancestral conserved features of a search against the eukaryotic MADBUB gene family. The consensus sequences of the detected conserved motifs are depicted as a sequence logo (colors reflect distinct amino acid properties and height of the letters indicates conservation of amino acids). Each features is assigned a differently colored shape. (B) Cartoon of the evolutionary scenario of 16 independent duplications of the MADBUB gene family throughout eukaryotic evolution, including a projection of conserved features onto the linear protein representation (on scale). Gene duplications are indicated by an arrow (red: high confidence, orange: uncertain). The subfunctionalized paralogs MAD and BUB are color brown and blue, respectively. Numbers indicate the clades in which the duplications occurred: {1}-mucorales; {2}- saccharomycetaceae; {3}-schizosaccharomycetes; {4}-pucciniomycetes: {5}-agaricomycetes (excluding early-branching species); {6}-vertebrates; {7}-teleost fish; {8}-nematodes; {9}-diptera (flies); {10}-albuginaceae (oomycete); {11}-ectocarpales (brow algae); {12}-aureococcus (harmful algae bloom); {13}-bryophytes (mosses); {14}-tracheophytes (vascular plants); {15}-magnoliaphytes (flowering plants); {16}-naegleria;

Figure 2 – Co-evolution of conserved features signify subfunctionalization of MAD and BUB after MADBUB duplication (A) Average clustering based on pearson distances of conserved ancestral feature correlation matrix (d=1-*r*) of MADBUB paralogs. Blue and red indicate co-presence or absence of features to be present in the

same paralog. (B) Evolutionary scenario of MADBUB subfunctionalization: MAD (cytosol) as a SAC effector and BUB (kinetochore) involved in SAC signal formation and kinetochore microtubule attachment.

Figure 3 – The evolutionary conserved cassette ABBA1-KEN2-ABBA2 in BUBR1 is essential for SAC signaling (A) Alignment of ABBA1-KEN2-ABBA2 cassette (red). Linkers (black) between ABBA motifs and KEN2 are indicated by {n}. The sequence logo on top is representative for all eukaryotic sequences (colors reflect distinct amino acid properties and height of the letters indicates conservation of amino acids). (B) Schematic representation of LAP-hBUBR1 mutants. Color-coding is consistent for each mutant in this figure (C) Immunoblots of BUBR1 and tubulin of mitotic lysates of HeLa FlpIn cell lines stably expressing LAP-tagged BUBR1 proteins. Cells were treated with siRNA (40nM) for 48 hours and cells were released and arrested into taxol after double thymidine block. (E) Time-lapse analysis of HeLa FlpIn cells expressing hBUBR1 mutants, treated with 20 μM STLC. Data (n=50, N=3) indicate the mean of cumulative fraction of cells that exit mitosis after nuclear envelope breakdown. Transparent regions represent the standard error of the mean. Values between accolades {} indicate the median value. Cells were score by cell morphology using DIC imaging (D); see for examples of SAC deficient (ΔABBA1/2) and proficient cells (wildtype). (F) Immunoblots of GFP, APC3 and CDC20 in LAP-BUBR1 precipitations (LAP-pulldown) in whole cell lysates of mitotic HeLa FlpIn cells stably expressing LAP-BUBR1 mutant constructs.

References

Alfieri, C., L. Chang, Z. Zhang, J. Yang, S. Maslen, M. Skehel, and D. Barford. 2016. Molecular basis of APC/C regulation by the spindle assembly checkpoint. *Nature*. 536:431–436. doi:10.1038/nature19083.

Andrews, P.D., Y. Ovechkina, N. Morrice, M. Wagenbach, K. Duncan, L. Wordeman, and J.R. Swedlow. 2004. Probing the catalytic functions of Bub1 kinase using the small molecule inhibitors BAY-320 and BAY-524. *Dev. Cell*. 6:253–268. doi:10.1016/S1534-5807(04)00025-5.

Bailey, T.L., M. Boden, F. a Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, and W.S. Noble. 2009. MEME S UITE: tools for motif discovery and searching. *Conflict*. 37:202–208. doi:10.1093/nar/gkp335.

Baron, A.P., C. von Schubert, F. Cubizolles, G. Siemeister, M. Hitchcock, A. Mengel, J. Schroder, A. Fernández-Montalván, F. von Nussbaum, D. Mumberg, and E.A. Nigg. 2016. Probing the catalytic functions of Bub1 kinase using the small molecule inhibitors BAY-320 and BAY-524. *Elife*. 5. doi:10.7554/eLife.12187.

Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268:78–94. doi:10.1006/jmbi.1997.0951.

Burton, J.L., and M.J. Solomon. 2007a. Mad3p, a pseudosubstrate inhibitor of APCCdc20 in the spindle assembly checkpoint. *Genes Dev.* 21:655–667. doi:10.1101/gad.1511107.

Burton, J.L., and M.J. Solomon. 2007b. Mad3p, a pseudosubstrate inhibitor of APCCdc20 in the spindle assembly checkpoint. *Genes Dev.* 21:655–67. doi:10.1101/gad.1511107.

Chao, W.C.H., K. Kulkarni, Z. Zhang, E.H. Kong, and D. Barford. 2012. Structure of the

427 mitotic checkpoint complex. *Nature*. 484:208–213. doi:10.1038/nature10896.

428 Crooks, G.E., G. Hon, J.-M.M. Chandonia, and S.E. Brenner. 2004. WebLogo: a sequence

429 logo generator. *Genome Res.* 14:1188–1190. doi:10.1101/gr.849004.

430 Darriba, D., G.L. Taboada, R. Doallo, and D. Posada. 2011. ProtTest 3: fast selection of

431 best-fit models of protein evolution. *Bioinformatics*. 27:1164–5.

432 doi:10.1093/bioinformatics/btr088.

433 Davey, N.E., J.L. Cowan, D.C. Shields, T.J. Gibson, M.J. Coldwell, and R.J. Edwards. 2012.

434 SLiMPrints: conservation-based discovery of functional motif fingerprints in

435 intrinsically disordered protein regions. *Nucleic Acids Res.* 40:10628–41.

436 doi:10.1093/nar/gks854.

437 Davey, N.E., and D.O. Morgan. 2016. Building a Regulatory Network with Short Linear

438 Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting Complex.

439 *Mol. Cell*. 64:12–23. doi:10.1016/j.molcel.2016.09.006.

440 DiFiore, B., N. Davey, A. Hagting, D. Izawa, J. Mansfeld, T. Gibson, and J. Pines. 2015. The

441 ABBA Motif binds APC/C activators and is shared by APC/C substrates and

442 regulators. *Dev. Cell*. 32:358–372. doi:10.1016/j.devcel.2015.01.003.

443 Eddy, S.R. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7.

444 doi:10.1371/journal.pcbi.1002195.

445 Elowe, S., K. Dulla, A. Uldschmid, X. Li, Z. Dou, and E.A. Nigg. 2010. Uncoupling of the

446 spindle-checkpoint and chromosome-congression functions of BubR1. *J. Cell Sci.*

447 123:84–94. doi:10.1242/jcs.056507.

448 Finn, R.D., J. Clements, and S.R. Eddy. 2011. HMMER web server: Interactive sequence

449 similarity searching. *Nucleic Acids Res.* 39:W29-37. doi:10.1093/nar/gkr367.

450 Izawa, D., and J. Pines. 2014. The mitotic checkpoint complex binds a second CDC20 to

451 inhibit active APC/C. *Nature*. 517:631–634. doi:10.1038/nature13911.

452 Kang, J., M. Yang, B. Li, W. Qi, C. Zhang, K.M. Shokat, D.R. Tomchick, M. Machius, and H.
453 Yu. 2008. Structure and Substrate Recruitment of the Human Spindle Checkpoint
454 Kinase Bub1. *Mol. Cell.* 32:394–405. doi:10.1016/j.molcel.2008.09.017.
455 Katoh, K., and D.M. Standley. 2013. MAFFT multiple sequence alignment software
456 version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
457 doi:10.1093/molbev/mst01.
458 Kawashima, S. a, Y. Yamagishi, T. Honda, K. Ishiguro, and Y. Watanabe. 2010.
459 Phosphorylation of H2A by Bub1 prevents chromosomal instability through
460 localizing shugoshin. *Science*. 327:172–177. doi:10.1126/science.1180189.
461 King, E.M.J., S.J.A. van der Sar, and K.G. Hardwick. 2007. Mad3 KEN Boxes Mediate both
462 Cdc20 and Mad3 Turnover, and Are Critical for the Spindle Checkpoint. *PLoS One*.
463 2:e342. doi:10.1371/journal.pone.0000342.
464 Klebig, C., D. Korinth, and P. Meraldi. 2009. Bub1 regulates chromosome segregation in a
465 kinetochore-independent manner. *J. Cell Biol.* 185:841–858.
466 doi:10.1083/jcb.200902128.
467 Kruse, T., G. Zhang, M.S.Y. Larsen, T. Lischetti, W. Streicher, T. Kragh Nielsen, S.P. Bjørn,
468 and J. Nilsson. 2013. Direct binding between BubR1 and B56-PP2A phosphatase
469 complexes regulate mitotic progression. *J. Cell Sci.* 126:1086–1092.
470 Lara-Gonzalez, P., M.I.F. Scott, M. Diez, O. Sen, and S.S. Taylor. 2011. BubR1 blocks
471 substrate recruitment to the APC/C in a KEN-box-dependent manner. *J. Cell Sci.*
472 124:4332–4345. doi:10.1242/jcs.094763.
473 Lischetti, T., G. Zhang, G.G. Sedgwick, V.M. Bolanos-Garcia, and J. Nilsson. 2014. The
474 internal Cdc20 binding site in BubR1 facilitates both spindle assembly checkpoint
475 signalling and silencing. *Nat. Commun.* 5:5563. doi:10.1038/ncomms6563.
476 London, N., and S. Biggins. 2014a. Signalling dynamics in the spindle checkpoint

477 response. *Nat. Rev. Mol. Cell Biol.* 15:736–47. doi:10.1038/nrm3888.

478 London, N., and S. Biggins. 2014b. Mad1 kinetochore recruitment by Mps1-mediated

479 phosphorylation of Bub1 signals the spindle checkpoint. *Genes Dev.* 28:140–152.

480 doi:10.1101/gad.233700.113.

481 Moyle, M.W., T. Kim, N. Hattersley, J. Espeut, D.K. Cheerambathur, K. Oegema, and A.

482 Desai. 2014. A Bub1-Mad1 interaction targets the Mad1-Mad2 complex to

483 unattached kinetochores to initiate the spindle checkpoint. *J. Cell Biol.* 204:647–

484 657. doi:10.1083/jcb.201311015.

485 Murray, A.W. 2012. Don't make me mad, Bub! *Dev. Cell.* 22:1123–5.

486 doi:10.1016/j.devcel.2012.05.020.

487 Murray, A.W., and D. Marks. 2001. Can sequencing shed light on cell cycling? *Nature*.

488 409:844–846. doi:10.1038/35057033.

489 Musacchio, A. 2015. The Molecular Biology of Spindle Assembly Checkpoint Signaling

490 Dynamics. *Curr. Biol.* 25:R1002–R1018. doi:10.1016/j.cub.2015.08.051.

491 Nguyen Ba, A.N., B.J. Yeh, D. van Dyk, A.R. Davidson, B.J. Andrews, E.L. Weiss, and A.M.

492 Moses. 2012. Proteome-Wide Discovery of Evolutionary Conserved Sequences in

493 Disordered Regions. *Sci. Signal.* 5:rs1-rs1.

494 Ogo, N., S. Oishi, K. Matsuno, J. Sawada, N. Fujii, and A. Asai. 2007. Synthesis and

495 biological evaluation of L-cysteine derivatives as mitotic kinesin Eg5 inhibitors.

496 *Bioorg. Med. Chem. Lett.* 17:3921–4. doi:10.1016/j.bmcl.2007.04.101.

497 Sacristan, C., and G.J.P.L. Kops. 2015. Joined at the hip: Kinetochores, microtubules, and

498 spindle assembly checkpoint signaling. *Trends Cell Biol.* 25:21–28.

499 doi:10.1016/j.tcb.2014.08.006.

500 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis

501 of large phylogenies. *Bioinformatics*. 30:1312–3.

doi:10.1093/bioinformatics/btu033.

Stanke, M., A. Tzvetkova, and B. Morgenstern. 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* 7 Suppl 1:S11.1-8. doi:10.1186/gb-2006-7-s1-s11.

Sudakin, V., G.K.T. Chan, and T.J. Yen. 2001. Checkpoint inhibition of the APC / C in HeLa cells. *Cell.* 154:925–936. doi:10.1083/jcb.200102093.

Suijkerbuijk, S.J.E., T.J.P. van Dam, G.E. Karagöz, E. von Castelmur, N.C. Hubner, A.M.S. Duarte, M. Vleugel, A. Perrakis, S.G.D. Rüdiger, B. Snel, and G.J.P.L. Kops. 2012a. The Vertebrate Mitotic Checkpoint Protein BUBR1 Is an Unusual Pseudokinase. *Dev. Cell.* 22:1321–1329. doi:10.1016/j.devcel.2012.03.009.

Suijkerbuijk, S.J.E., M.H.J. Van Osch, F.L. Bos, S. Hanks, N. Rahman, and G.J.P.L. Kops. 2010. Molecular causes for BUBR1 dysfunction in the human cancer predisposition syndrome mosaic variegated aneuploidy. *Cancer Res.* 70:4891–4900. doi:10.1158/0008-5472.CAN-09-4319.

Suijkerbuijk, S.J.E., M. Vleugel, A. Teixeira, and G.J.P.L. Kops. 2012b. Integration of Kinase and Phosphatase Activities by BUBR1 Ensures Formation of Stable Kinetochore-Microtubule Attachments. *Dev. Cell.* 23:745–755. doi:10.1016/j.devcel.2012.09.005.

Tang, Z., R. Bharadwaj, B. Li, and H. Yu. 2001. Mad2-Independent inhibition of APC Cdc20 by the mitotic checkpoint protein BubR1. *Dev Cell.* 1:227–237. doi:10.1016/S1534-5807(01)00019-3.

Taylor, S.S., E. Ha, and F. McKeon. 1998. The human homologue of Bub3 is required for kinetochore localization of Bub1 and a Mad3/Bub1-related protein kinase. *J. Cell Biol.* 142:1–11. doi:10.1083/jcb.142.1.1.

Tromer, E., B. Snel, and G.J.P.L. Kops. 2015. Widespread recurrent patterns of rapid repeat evolution in the kinetochore scaffold KNL1. *Genome Biol. Evol.* 7:2383–2393.

527 doi:10.1093/gbe/evv140.

528 Vleugel, M., T. a. Hoek, E. Tromer, T. Sliedrecht, V. Groenewold, M. Omerzu, and G.J.P.L.

529 Kops. 2015. Dissecting the roles of human BUB1 in the spindle assembly

530 checkpoint. *J. Cell Sci.* 128:2975–2982. doi:10.1242/jcs.169821.

531 Vleugel, M., E. Hoogendoorn, B. Snel, and G.J.P.L. Kops. 2012. Evolution and Function of

532 the Mitotic Checkpoint. *Dev. Cell.* 23:239–250. doi:10.1016/j.devcel.2012.06.013.

533 Wu, J., S. Kasif, and C. DeLisi. 2003. Identification of functional links between genes

534 using phylogenetic profiles. *Bioinformatics*. 19:1524–30.

535 doi:10.1093/BIOINFORMATICS/BTG187.

536 Xu, P., E. a Raetz, M. Kitagawa, D.M. Virshup, and S.H. Lee. 2013. BUBR1 recruits PP2A

537 via the B56 family of targeting subunits to promote chromosome congression. *Biol.*

538 *Open*. 2:479–86. doi:10.1242/bio.20134051.

539 Yamaguchi, M., R. VanderLinden, F. Weissmann, R. Qiao, P. Dube, N.G. Brown, D.

540 Haselbach, W. Zhang, S.S. Sidhu, J.-M. Peters, H. Stark, and B.A. Schulman. 2016.

541 Cryo-EM of Mitotic Checkpoint Complex-Bound APC/C Reveals Reciprocal and

542 Conformational Regulation of Ubiquitin Ligation. *Mol. Cell.* 63:593–607.

543 doi:10.1016/j.molcel.2016.07.003.

544

Figure 1

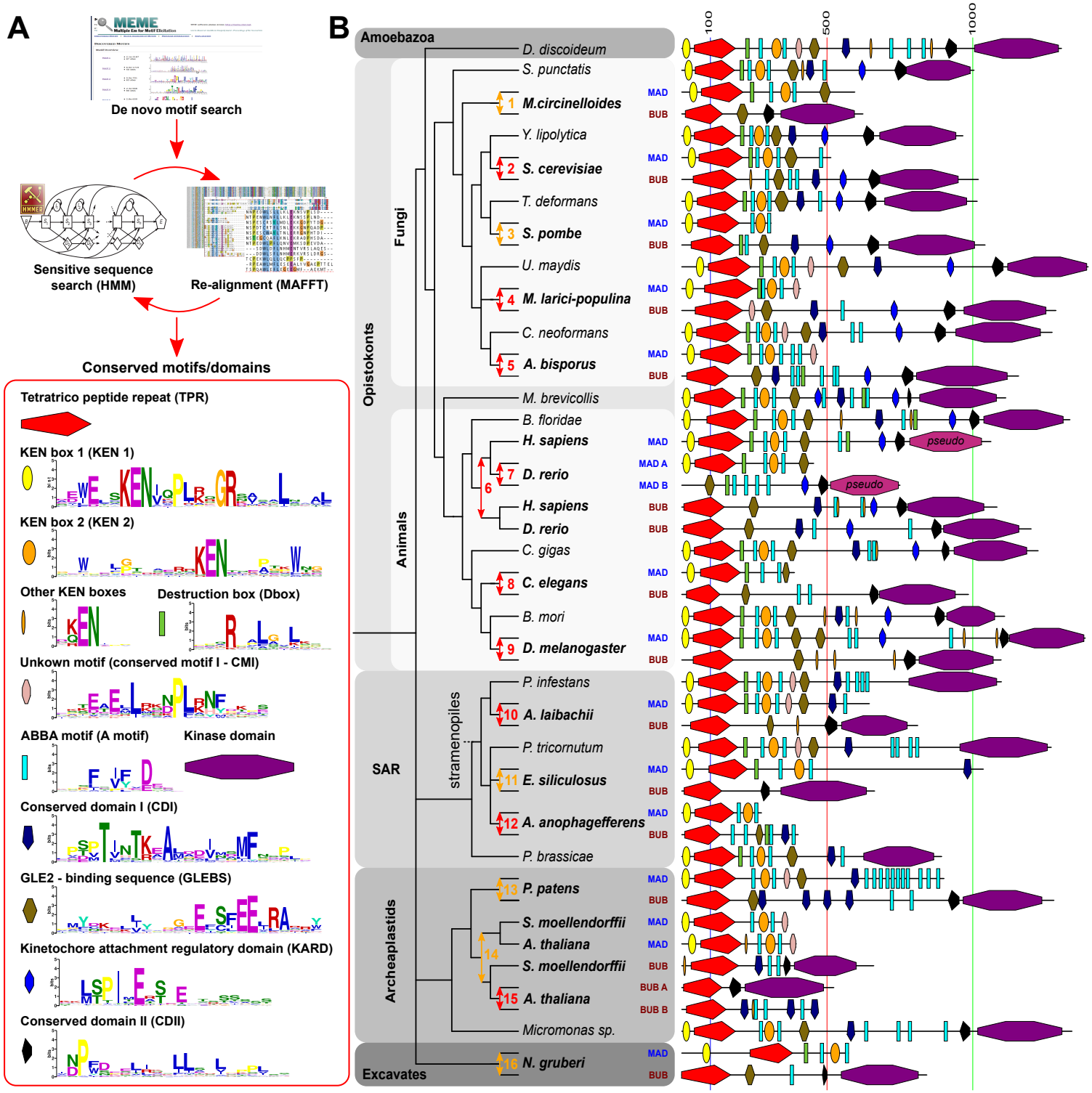
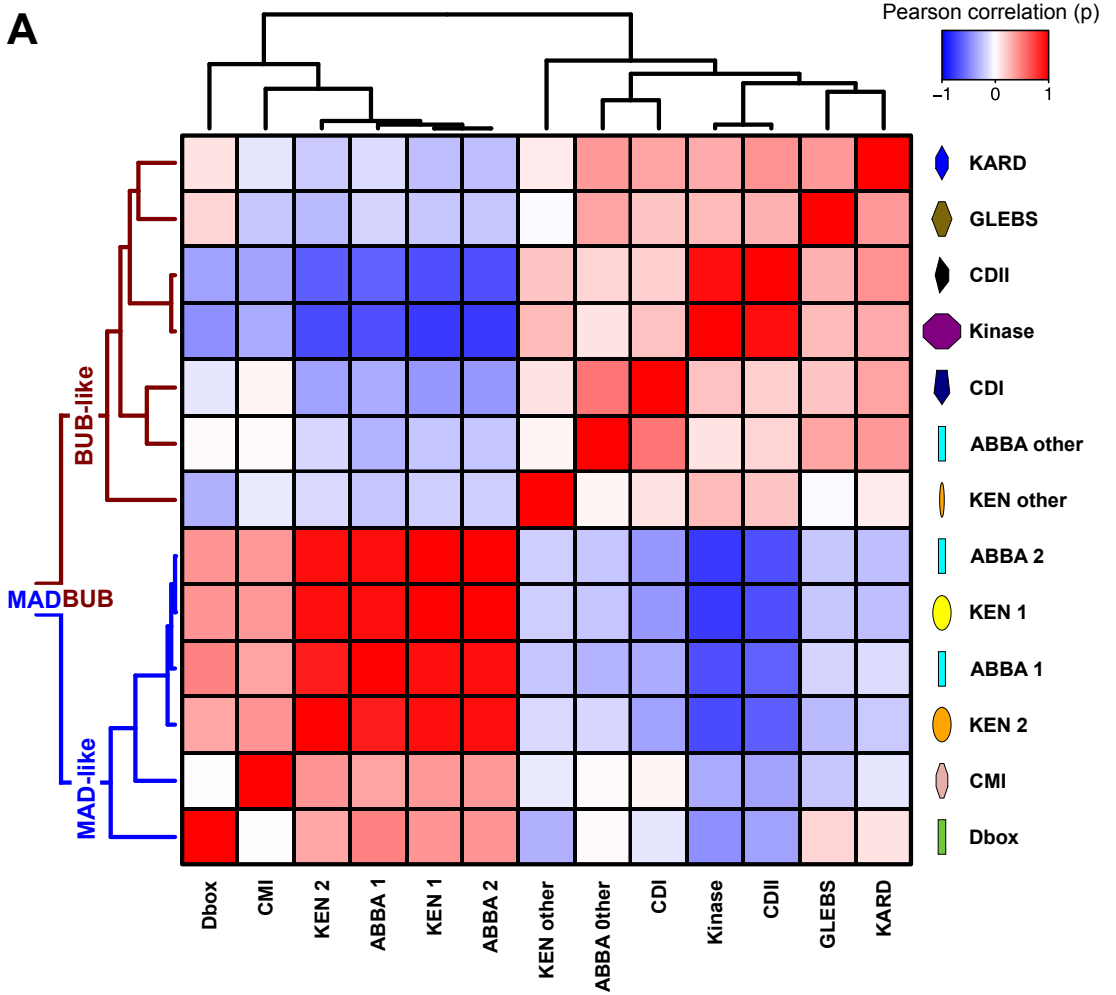


Figure 2 Tromer et al. 2016



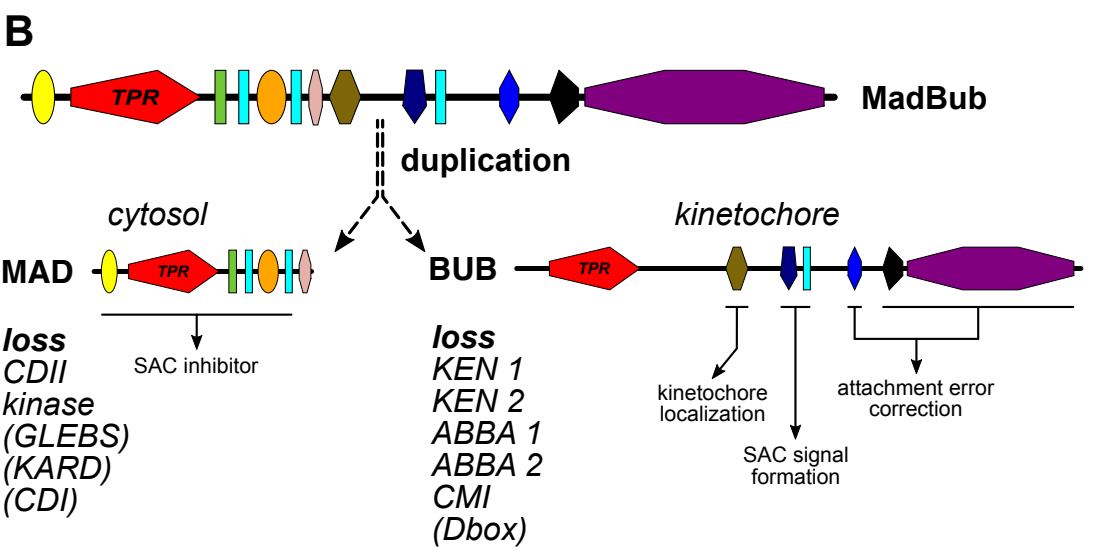


Figure 3 Tromer et al. 2016