

Supplementary information

Supplementary Figure 1. Initial algorithm for prioritisation of Discovery Phase variants for further evaluation. Number of genes or variants remaining after each stage are shown.

Exclude variants outside protein-coding regions and splice sites and synonymous variants

▼ N=13,829 variants

Exclude variants in 1000 genomes or dbSNP at frequency ≥ 0.001

▼ N=5,569 variants

Exclude variants present in 202 patients without cancer in WGS500

▼ N=5,228 variants

Test whether any genes with predicted strongly deleterious variants present in ≥ 4 families

▼ N=23 genes

Exclude duplicated sequences, mapping errors and poor quality reads not already detected

N=0 genes

Supplementary Figure 2a. Clustal omega multi-sequence alignments of POLE across species

Residue 424 in humans is highlighted.

CLUSTAL O(1.1.0) multiple sequence alignment

```

Homo      -----
Mus       -----
Anolis    MVAIPGPCREPIGSLRQPTKSWLNSNSQATKAQAPTLRNASEMSHTGSPLPGSVFQEGRQ
Xenopus   -----
Danio     -----
Drosophila -----
Caenorhabditis -----
Saccharomyces -----
Schizosaccharomyces -----
    
```

```

Homo      -----
Mus       -----
Anolis    PLQNTLQPIQGRGAGLGALNGRGRSRGWAGPSLFFVGRSRWRPAEGVWPLPKGDRRKSGRGF
Xenopus   -----
Danio     -----
Drosophila -----
Caenorhabditis -----
Saccharomyces -----
Schizosaccharomyces -----
    
```

```

Homo      -----MSLRSGGRRR-----
Mus       -----MVLNRNSGRRH-----
Anolis    PGLSVRLRGRNVIGPGHVVEAAVPRAGGVSRGFPSARESEGVSEMAARRGGRRWRGEE
Xenopus   -----MVLQNSGKLR-----
Danio     -----MVLQNSGRYK-----
Drosophila -----MSDSSKGKVLQNTGKFVS-----
Caenorhabditis -----
Saccharomyces -----MMFGK--KKNNGGSSTARYSAGNK-----
Schizosaccharomyces -----MPLKTARGASKYQF--RKFNGNYNGKSKSNGR-----
    
```

```

Homo      ADPGADGEASRDDGATSSVSALKRRLERSQWTDKMDLRFGFERLKEP-----
Mus       PEPGADGEGSRDDGPPSSVSALKRRLERSQWTDKMDLRFGFERLKEP-----
Anolis    APPGEGPGEAPKGGDDGSSLSALKRRLERSHWTDTIDARFGFERIKEP-----
Xenopus   AEPRG--DGENQQDDASASSAAKRLERSQWTDKIDAQYGFERIREP-----
Danio     ADRGG--DQDNQQDDASALSAVKRRLERSQFTDSMDERFGFERMKEP-----
Drosophila -----ENRTEGDDFFNEAGYRQSRENDKIDSKYGFDRVKDS-----
Caenorhabditis -----MSSKDDILAQA VENDSNYKERLALIRSNDEIDAKLGF SRYTGL-----
Saccharomyces -----YNTLSNNYALS AQQLLNASKIDDIDSMMGFERYVPPQYNGRFD AKDIDQI-----
Schizosaccharomyces -----FAKSTEEVGFN-DPMKIVYKKN EIDRMMGFDSYEGG-----
    
```

. : :* **.

```

Homo      GEKTGWLINMHPTEILDE-----DKR-LGSAVDYFYIQDDGSRFKVA
Mus       GERTGWLINMHPTEILDE-----DKR-LVSAVDYFYIQDDGSRFKVA
Anolis    TERTGWLINMHPTEILDE-----DRH-LVSAVDYFYIQEDGSRFKVA
Xenopus   GEKTGWLLNMHPNEILDE-----DKR-LVSAVDYFYIQEDGSRFKVA
Danio     GEKTGWLINMHPTEILDD-----DKR-MISAVDYFYIQEDGNRFKVA
Drosophila QERTGYLINMHSNEVLDE-----DRR-LIAALDLFFIQMDGSRFKCT
Caenorhabditis QEKKGFLINIQPSELVDE-----QTKVIISVVDYFFISDMDERFKIS
Saccharomyces PGRVWGWLTMHATLVSQETLSSGSGNGGNSNDGERVTTNQGISGVDFYFLDEEGGSFKST
Schizosaccharomyces QPREAWLLNVHPTVIES-----TKGNSTLSAVDFYFIQDDGDTFRCT
    
```

: .* *: . : . : : * :* . * : :

```

Homo      LPYKPYFYIATRKG CER-EVSSFLSKKFQGKIAKVETVPKEDLDLPNHLVGLKRN YIRLS
Mus       LPYMPYFYIAARKG CDR-EVSSFLSKKFQGKIAKLENVPKEDLDLPNHLVGLKRSYIKLS
Anolis    LPYKPYFYVAAQKG CDR-EVSSFLSKKFQGKIAKLETVPKEDLDLPNHLVGLKRN YLKLS
Xenopus   LPYKPYFYVATKKG CER-EVSSFLSKKFQGKLAKLETIPKEDLDLPNHLVGLKRGYIKLS
Danio     LPYKPYFYIATKKNTER-EVISFLSKKFQGVAKLESVPKEDLDLPNHLVGLKRN YIKLS
    
```

Drosophila VAYQPYLLIRPEDNMHL-EVARFLGRKYSQISGLEHITKEDLDLPNHLGLQQQYIKLS
Caenorhabditis YPFRPYFYIATLDGFEF-QVSSYLSKKYGAQ-TAVEHMDREDLDLKDHLGLKTTYIKLS
Saccharomyces VVYDPYFFIACNDESRVNDVEELVKKYLESCLKSLQIRKEDLTMNHLGLQKTLIKLS
Schizosaccharomyces IPYSPYFYIAAREGKEA-LVDDYLKKKFEVGLIKSTTRIFKEDLQKNHIVGYQKLYIKLV
: **: : . * : : . : :*** :*: * : : :*

Homo FHTVEDLVKVRKEISPAVKKNREQDHASDAYTALLSSVLQRGGVITD-EEETSKKIADQL
Mus FHTVEDLVKVRKEISPAVKKNREQDHASDEYTTMLSSILQGGSVITD-EDETSKKIADQL
Anolis FNSVDDLKVRKEISPAVRKNQERDQSVDAYTTMLASALAGGSLSSK-DKEPSKKMSNQM
Xenopus FHSVDDLKVRKEISPAVRKNKERDHASDTYTAMLSSALTGNMGAEEEGPSKKISDQM
Danio FNTVDDLKVRKEISPAVRKNREREKSNDAYTSMSSALVGGSVVTEDEGGSSKKMTEQYL
Drosophila FLNQTAMTKVRRELMSAVKRNQERQKSNITYMQMLATSLAQSSAGSE-DATLGRQQDYM
Caenorhabditis FTSTVELIKIRKELMPLVRKNTDRIKKESAYADYLARNLSGKGGSDKQ----QLNGDIL
Saccharomyces FVNSNQLFEARKLLRPILQDNANNNVQRNIYNVA----ANG-----SEKVDK
Schizosaccharomyces FDNLNDLQAVRKSLSAVKANSSQQDAVDAYTNLSSENLNGIIE-----NAFEDPL
* . : : : : : * . * . : . :

Homo DNIVDMREYDVPYHIRLSIDLKIHVAHWYNVRYRGNAPVEITRRDDLVERPDPVVLAFD
Mus DNIVDMREYDVPYHIRLSIDLRIHVAHWYNVFRGNAPVEITRRDDLVERPDPVVLAFD
Anolis DNIVDMREYDVPYHIRLSIDLKIHVAHWYNVRYRGSTFPPEITRRDDLVERPDPVVLAFD
Xenopus ENIVDMREYDVPYHVRVSDIDLKIHVAHWYNIRYRGSPPPEITRRDDLVERPDPVVLAFD
Danio DNILDMREYDVPYHVRVSDIDLKIHVAHWYNVRYRGSAYPPEIVRRDDLVERPDPVVLAFD
Drosophila DCIVDIREHDVPYHVRVSDILRIFCGQWYNIRCRSGVELPTITCRPDILDRPEPVLAFD
Caenorhabditis NQIVDIREYDVPFHMVRSIDEKIFVGLWYDVKGIGPNRVPTIKRDLPLFHAKPKVLAFD
Saccharomyces HLIEDIREYDVPYHVRVSDIKDIRVGKWKYVTTQQGF-----IEDTRKIAFADPVMMAFD
Schizosaccharomyces NHVLDIREYDVPYHSRTLIDLNIRVGQWYTVTSYHEGHVQ--ISLLASRIERAEPIMAFD
. : *:*:*:*:*:* * * * * * . * * : : . * : :***

Homo IETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSEDIEDFEFTPKPEYEGPFCVF
Mus IETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSEDIEDFEFTPKPEYEGPFCVF
Anolis IETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSEDIEDFEFTPKPEYEGPFCVF
Xenopus IETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSEDIEDFEFTPKPEYEGPFCIF
Danio IETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSEDIEDFEFTPKPEYEGPFTIF
Drosophila IETTKLPLKFPDAQTDQVMMSYMIDGQGYLITNREIISNVDDFEYTPKPEFEGNFIVF
Caenorhabditis IETTKLPLKFPDRESDEIMMISYMVDGRGFLINREIVSADINAFEYTPKAEYIGEFTVW
Saccharomyces IETTKPPLKFPDSAVDQIMMISYMIDGEGFLITNREIISEDIEDFEYTPKPEYPGFFTIF
Schizosaccharomyces IETTKLPLKFPDSSFDKIMMISYMIDGQGYLITNREIISQNIEDFHYTPREEFEGPFIIF
***** * * * * * : : * * * * * : * : * : * : * : * : *

Homo NEPDEAHLIQRWFEHVQETKPTIMVTYNGDFFDWPVFEARAAVHGLSMQOEIGFQKDSQG
Mus NEPDEVHLIQRWFEHIQETKPTIMVTYNGDFFDWPVFEARAAIHGLSMYQEIGFQKDSQG
Anolis NEPDEAHLIQRWFEHVQDTKPTIMVTYNGDFFDWPVFEARAAHGMSMQKEIGFQKDNQG
Xenopus NEPDEAHLIQRWFEHMQETKPNIVTYNGDFFDWPVFEARAAVHGLSMQOEIGFQKDNQG
Danio NEPDEASLIQRWFEHVHETKPNIVTYNGDFFDWPVFEARAAQLGLSMHREIGFQKDNQG
Drosophila NEENEMQLLQRFDDHIMEVVRPHIVTYNGDFFDWPVFEARAAVYLDLDMKQIEIGFQKDRG
Caenorhabditis NEKDEAALIRKFFDHILQVRPNIVTYNGDFFDWPVFEARAKIRGNMEREIGFSKDSAD
Saccharomyces NENDEVALLRFFFEHIRDVRPTVISTFNGDFFDWPVFIHNRKIHGLDMFDEIGFAPDAEG
Schizosaccharomyces NEPDEVGLLQRFFFKHIRSAKPSVIVTYNGDFFDWPVFDARAAFHGLNLTEETGFFRDAED
** :* * : : : * . . . : * : . * : * * * * * : . * : : : * * *

Homo EYKAPQCIHMDCLRWVKRDSYLPVGSNHLKAAAKAKLGYDPVELDPEDMCRMATEQPQTL
Mus EYKAPQCIHMDCLRWVKRDSYLPVGSNHLKAAAKAKLGYDPVELDPEDMCRMATEQPQTL
Anolis EYKSSQCIHMDCLRWVKRDSYLPVGSNHLKAAAKAKLGYDPVELDPEEMCRMATEEPQTL
Xenopus EYKSPPCIHMDCLRWVKRDSYLPVGSNHLKAAAKAKLGYDPVELDPEEMCRMATEEPQVL
Danio EYKASQAIHMDCLRWVKRDSYLPVGSNHLKAAAKAKLGYDPVELDPEEMCRMATEEPQTL
Drosophila NYLSRPAIHMDCLCWVKRDSYLPVGSQGLKAVAKAKLRYDPVELDPEDMCRMATEEPQVL
Caenorhabditis EYKSRNCIHMDAFAFRWVKRDSYLPVGSQNLKAVTKAKLRYDPVEVEPELMCKMAREQPQQL
Saccharomyces EYKSSYCSHMDCFRWVKRDSYLPVGSQGLKAVTQSKLGNPIELDPPELMTPYAFKPKQHL
Schizosaccharomyces EYKSSYCSHMDCFRWVKRDSYLPVGSQGLKAVTVSKLGNPIELDPPELMTPYASEKPKQVL
:* : . * * * : * * * * * * * : * * * : : * * * : * * * * * * * * * * *

Homo ATYSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEVLRKGSGLTCEALLMVQAFHANI I
Mus ATYSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEVLRKGSGLTCEALLMVQAFHANI I
Anolis ATYSVSDAVATYYMYMKYVHPFIFALCTIIPMEPDEVLRKGSGLTCEALLMVQAYHANI I
Xenopus ATYSVSDAVATYYMYMKYVHPFIFALCTIIPMEPDEVLRKGSGLTCEALLMVQAYHANI I
Danio ATYSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEVLRKGSGLTCEALLMVQAYHNI I
Drosophila ANYSVSDAVATYYLYMKYVHPFIFALNTIIPMEPDEILRKGSGTLTCELLMVEAYHAQIV

Caenorhabditis ANYSVSDAVSTYYLYMKYVHQFIFALCTIIPLGADDVLRKGSGLTCEALLMVEAFHNNIV
Saccharomyces SEYSVSDAVATYYLYMKYVHPFIFSLCTIIPLNPDVTLRKGVTGLCEMLLMVQAYQHNIL
Schizosaccharomyces AQYSVSDAVATYFLYMKYVHPFIFSLCNIIPLNPDVTLRKGVTGLCETLLTVEACTKNI
: *****: : : ***** * : * . * : * : * : * : * : * : * : *

Homo FPNKQEQ-EFNKLTDDGHVLDSETYVGGHVEALESVFRSDIPCRFRMNPAADFLLQRV
Mus FPNKQEQ-EFNKLTDDGHMLDAETVYVGGHVEALESVFRSDIPCRFRMNPAADFLLQRV
Anolis FPNKQEQ-EFNKLTEDGHVLDSETYVGGHVEALESVFRSDIPCRFRMNPAADFLLERV
Xenopus FPNKQEA-VFNKLTSDGHVLDSETYVGGHVEALESVFRSDIPCRFRMNPAADFLLQRV
Danio FPNKQEQ-VFNKLTDDGHVLDSETYVGGHVEALESVFRSDIPCRFRMNPAADFLLQRV
Drosophila YPNKHQS-ELNKLSDGHVLDSETYVGGHVEALESVFRSDIPCRFRMLDPAVQKQEQV
Caenorhabditis FPNKYTGPEETRFKSDGHRVSESYVGGHVEALEAGVFRADIPAKFRSLVPALEQLKSEI
Saccharomyces LPNKHTDPIER--FYDGHLLESETYVGGHVESLEAGVFRSDLKNEFKIDPSAIDELLQEL
Schizosaccharomyces LPNKHVDASQK--FFDGHLLESETYVGGHVESLESVFRSDLPTNFNMDPKVYEELILQL
*** : * : : ***** : * : * : * : * : * : * : *

Homo EKTLRHALEEEEEKVPVEQVTNFEEVCDEIKSKLASLKDVPRIECLPIYHLDVGMYPNI
Mus EKTMRHAIEEEEEKVPVEQATNFQEVCEQIKTKLTSCLKDVPRIECLPIYHLDVGMYPNI
Anolis EKTLRHAIEVEEGIPLDQVTNFQEVCEDEIKVKLRSCLKDVPRIECLPIYHLDVGMYPNI
Xenopus ELTLRHAIEEEEEKIPLEQVTNFQEVCEEIKKKLNSLKEVPRIECLPIYHLDVGMYPNI
Danio ERTLRAIEEEEEKIPLEQVTNFNEVCDEIKRKLISLKEVPRIECLPIYHLDVGMYPNI
Drosophila DAVLRHAIEVEEGIPLEKVLNLDEVRQEIYVQGLQGLHDIIPNRLEQPIYHLDVGMYPNI
Caenorhabditis QETLRKELAREFEVTLDDQVDFDEQCAEVQDAFDGMINVPTRELPRIYHLDVGMYPNI
Saccharomyces PEALKFSVEVENKSSVDKVTNFEEIKNQITQKLELKENNIRNELPLIYHVDVASMYPNI
Schizosaccharomyces DKALDFSLTVENNVNVDIEIENYEEVRDSILKKLSLDLDRPKRSEKPRIYHLDVASMYPNI
. : : * : : : * : : : * * * * * : * : * * * * *

Homo ILTNRLQPSAMVDEATCAACDFNKP GANCQRKMAWQWRGEFMPASRSEYHRIQHOLESEK
Mus ILTNRLQPSAIVDEATCAACDFNKP GASCQRKMAWQWRGEFMPASRSEYHRIQHOLESEK
Anolis ILTNRLQPSAMVDEASCAACDFNKP GANCQRRTWQWRGEFMPASRSEYHRIQQOLESEK
Xenopus ILTNRLQPSAMVDEVTCACDFNKP GATCQRQMTWQWRGEFMPASRSEYHRIQQOLESEK
Danio ILTNRLQPSAMVDEATCAACDFNKP GANCQRRTWQWRGEIMPASRSEFHRIQQOLESEK
Drosophila ILTNRLQPSAMVSDLDCAACDFNKP GVRCKRSMDWLWRGEMLPASRNEFQRIQQOLETEK
Caenorhabditis ILTNRLQPCAMVTEEICMGCSYNKPDACEKRTMAWEWRGELTPATRGEYQQIMQQLEAES
Saccharomyces MTTNRLQPDSIKAERDCASCFNRP GKTCAKRLKAWWRGEFFPSKMDEYNMIKRALQNET
Schizosaccharomyces MITNRLQPDSVKDESFCATCDLNVPNKTCDRRMVWAWRGEYYPAKKGEYHMIYSALQSER
: * * * * * : : : * * * * * : * * * * * : * * * * * : * * * * *

Homo FPPLFPEGPA--RAFHELSTREEQAKYEKRLADYCRKAYKKIHTKVEERLTTICQRENS
Mus FPPLFPEGPA--RAFHELSTREEQAKYEKRLADYCRKAYKKIHVTKVEERLTTICQRENS
Anolis FPPLFPEGPA--RAFHELSTREEQAKYEKRLADYCRKAYKKIHTKVEERLTTICQRENS
Xenopus FPPLFPEGPA--RAFHELSTREEQAKYEKRLADYCRKAYKKIHVTTRLEEKVTTICQRENS
Danio FPPVFPNGPP--RAFHELSTREEQAKYEKRLADYCRKAYKKVHLTRLEERLTTICQRENS
Drosophila FPPLFPEGPA--RAFHELSTREEQAKYEKRLADYCRKAYKKIHTKVEERLTTICQRENS
Caenorhabditis F-----GKPP--KHFHMLERSEREA IEMKRIKDYSRVYKTHLTRLEMRETTICQRENH
Saccharomyces FPNKKNFSSKVVLTDFDELSYADQVIHKKRLTEYSRKVYHRVYKVEIVEREATVCQRENH
Schizosaccharomyces FPGPTPFSPF--RSFQELSPSEQAAMVQKRIADYSRKVIYHRLYDNTVIERETIICQRENS
* * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Homo FYVDTVAFRRDRRYEFKGLHKVWKKKLSAAVEVGDAAEVKRCNMEVLYDSLQLAHKC I L
Mus FYVDTVAFRRDRRYEFKGLHKVWKKKLSAAVEVGDASEVKRCNME ILYDSLQLAHKC I L
Anolis FYVDTVAFRRDRRYEFKGLHKVWKKKLSAAVEVGDASEVKRCNME ILYDSLQLAHKC I L
Xenopus FYVDTVAFRRDRRYEFKGLHKVWKKKLSAAVEVGDAAEIKRCNME ILYDSLQLAHKC I L
Danio FYVDTVAFRRDRRYEFKGLHKVWKKKLSAAVEVGDAAEIKRCNME ILYDSLQLAHKC I L
Drosophila FYVDTVAFRRDRRYEFKGLTKVAKASVNAAVASGDAAEIKAAKGREVLYDSLQLAHKC I L
Caenorhabditis FYVETVKAFRDRRYEYKDMKKAKGRFDQAQATNDLATMTTSKLEMVLYDSLQLAHKC I L
Saccharomyces FYVDTVKSFRDRRYEFKGLAKTWKGNLSKIDPSDK-HARDEAKMIVLYDSLQLAHKV I L
Schizosaccharomyces FYIDTVKSFRDRRYEFKGLQKQWVQVLAALKEKGGGLAEIEEAKMIVLYDSLQLAHKV I L
* * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Homo NSFYGYVMRKARWYSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNSF
Mus NSFYGYVMRKARWYSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNSF
Anolis NSFYGYVMRKARWYSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNSF
Xenopus NSFYGYVMRKARWYSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNSF
Danio NSFYGYVMRKARWYSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNTF
Drosophila NSFYGYVMRRGARWHSMPMAGIVCLTGSNIITKAREI IERVGRPLELDTDGIWCVLPNSF
Caenorhabditis NSFYGYVMRKGRWFSMEMAGIVCF TGANIITQARELIEQIGRPLELDTDGIWCVLPNSF

Saccharomyces NSFYGYVMRKGSRWYSMEMAGITCLTGATIIQMARALVERVGRPLELDTDGIWCILPKSF
Schizosaccharomyces NSFYGYVMRKGSRWYSIEMAGITCLTGATIIQMARQIVERAGRPLELDTDGIWCILPESF
*****:*.**.*: ****.* **:.** ** :*: * *****:*.**.*

Homo PENFVFKTTNVKKPKVTISYPGAMLNIMVKEGFTNDQYQELAEPS---SLTYVTRSENSI
Mus PENFVIKTTNAKKPKLTISYPGAMLNIMVKEGFTNHQYQELTEPS---SLTYVTHSENSI
Anolis PENFVIKSTHTKKPKVTISYPGAMLNIMVKEGFTNDQYQELVDPA---SLQYVCRSENSI
Xenopus PENFVIKSTNAKKPKVTISYPGAMLNILVKEGFTNHQYQELSDPA---TLTYITRSENSI
Danio PENFVIKSTNEKKPKVTISYPGAMLNIMVKEGFTNHQYQELVDAA---SLTYETRAENSI
Drosophila PQEFTIHTSHEKKKKINISYPNAVLNMTMVKDHFNDQYHELKRDKENNLPKYDIRDENS
Caenorhabditis PENVTFKLKNHKRSSVTVSYPGAMLNALVYEGFTNHQYHTLEKDG-----SYKSSSENSI
Saccharomyces PETYFFTLLENGK--KVLISYPCSMNLNVRVHVKFTNHQYQELKDPDPL---NYIYETHSENTI
Schizosaccharomyces PENFEFKKSGG--KVFISSYPCVMLNHLVHEKFTNHQYSALKDPE---KLVYETTSSENSI
: : :: ** * : **.* * . * **.*

Homo FFEVDGPYLAMILPASKEEGKKLKRYAVFNEDGSLAELKGFVKKRREGELQLIKIFQSSV
Mus FFEVDGPYLAMILPASKEEGKKLKRYAVFNEDGSLAELKGFVKKRREGELQLIKIFQSSV
Anolis FFEVDGPYLAMILPASKEEGKKLKRYAVFNEDGSLAELKGFVKKRREGELQLVKIFQSSV
Xenopus FFEVDGPYLAMILPASKEEGKKLKRYAVFNEDGSLAELKGFVKKRREGELQLIKIFQSSV
Danio FFEVDGPYLAMILPASKEEGKKLKRYAVFNEDGSLAELKGFVKKRREGELQLIKIFQSSV
Drosophila FFEVDGPYLAMVLPAAKEEGKKLKRYAVFNFDGTLAELKGFVKKRREGELQLIKNFQSSV
Caenorhabditis YFEVDGPYQCMILPASKEEGKKLKRYAVFNLDGSLAEMKGFELKRRGELNI IKHFQGCV
Saccharomyces FFEVDGPYKAMILPSSKEEGKGIKKRYAVFNEDGSLAELKGFELKRRGELQLIKNFQSDI
Schizosaccharomyces FFEVDGPYRAMILPASTEEGKLNKRYAVFNFDGSLAELKGFVKKRREGELKLIKDFQSQI
:***** .*:**:.**** :***** *:***:****:*****:.* ** . :

Homo FEAFKLGSTLEEVYGSVAKVADYWLDVLYSKAANMPDSELFELISENRSMSRKLEDYGEQ
Mus FEAFKLGSTLEEVYGSVAKVADYWLDVLYSKAANMPDSELFELISENRSMSRKLEDYGEQ
Anolis FEAFKLGSTLEEVYAAVAKVADYWLDVLYSKAANMPDSELFELISENRSMSRKLEDYGEQ
Xenopus FEAFKLGSTLEEVYASVAKVADYWLDVLYSKAANMPDAELFELISENRSMSRRLEDYGEQ
Danio FEAFKLGSTLEEVYASVAKVADYWLDVLYSKAANMPDAELFELISENRSMSRKLEDYGEQ
Drosophila FEAFLAGSTLEECYASVAKVADYWLDVLYSRGSLNLPDSELFELISENKSMSKLEEYGAQ
Caenorhabditis FKTFLNGKTLEETYKAVAADADHWLDILHSHGADLTDEELFDLISENRSMSRKLEDYGAQ
Saccharomyces FKVFLEGDTLEGCYASAVASVCNRWLDVLDHGLMLEDEDLVSLICENRSMSKTLKEYEGQ
Schizosaccharomyces FKVFLKGSLSLEECYQEVAYVADTWLEILFTKGSNLTDDELIELISENRSMSKALSEYGSQ
*:.** * :** * ** .: **:* :. : * :*.**.*:****: *:.** *

Homo KSTSISTAKRLAEFLGDQMVKDAGLSCRYIISRKPEGSPVTERAIPLAIFQAEPTVRKHF
Mus KSTSISTAKRLAEFLGDQMVKDAGLSCRYIISRKPEGSPVTERAIPLAIFQAEPTVRKHF
Anolis KSTSISTAKRLAEFLGDQMVKDAGLSCRFIISKKPEGAPVTERAIPLAIFQAEPSVRRHY
Xenopus KSTSISTAKRLAEFLGDQMVKDAGLSCRYVISRKPEGSPVTERAIPLAIFQAEVGVKRHY
Danio KSTSISTAKRLAEFLGDQMVKDAGLSCRYVISRKPEGSPVTERAIPLAIFQAEESVKKHF
Drosophila KSTSISTAKRLAEFLGDMVKDAGLACKYIISKKPEGAPVTERAIPLAIFQSEPSVRRHH
Caenorhabditis KSTSISTAKRLAEFLGDDMVKDAGLACMFIISKHPGAPVTERAIPVAIFKSDAKVRSHY
Saccharomyces KSTSITTARRLGDFLGEDMVKDGLQCKYIISKPFNAPVTERAIPVAIFSDIPIKRSF
Schizosaccharomyces KSTSITTARRLADFLGDQMTKDKGLACRFIISASPKGRPVAERAVPVAIFFAEESVRRHF
*****:*.**.*:****:*.**.* ** * :*: * **.*:*.**.*:****: *:.** *

Homo LRKWLKSSSL-QDFDIRAILDWDYIYERLGSQAIQKIITIPAALQQVKNPVPRVKHPDWLH
Mus LRKWLKSSSL-QDFDIRTILDWDYIYERLGSQAIQKIITIPAALQQVKNPVPRVKHPDWLH
Anolis LRKWLKSPSL-QDLDIRTILDWDYIYERLGNITQKIITIPAALQQVKNPVPRVHHPDWLH
Xenopus LRKWLKNPSL-QDLDIRSILDWDYIYERLGSQAIQKIITIPAALQQVKNPVPRVHHPDWLH
Danio LRKWLKMPSL-HDLDIRSILDWSYIYERLGSQAIQKIITIPAALQQVKNPVPRVHHPDWLH
Drosophila LRRWLKDNTM-GDADIRDVLDWNYIYERLGGTIQKIITIPAALQGLANPVPRVQHPDWLH
Caenorhabditis IRKWTQVDFNEDTDIRDMLDWDYIYERLFGSCIQKIITIPAALQGISNPVPRVPHDPWLQ
Saccharomyces LRRWTLDPSSL-EDLDIRTIDWGYIYERLGSQAIQKIITIPAALQGVSNPVPRVEHPDWLQ
Schizosaccharomyces LRLWLKDNGL-YDVIDIRIDWYIYERLGSQVQKLISIPAALQRIISNPVTRFPLPDWLQ
:* * : * ** :*: ** * :*: * :*: * :*: * :*: * :*: *

Homo KKLEKNDVYKQKKISELFTLEGRQVTMA--EASEDSRPSAPDMEDFLV--KLPHPA
Mus KKLEKNDIYKQKKISELFLVLEGRQIVMA--QASENSLSLCTPDMEDIGLT--KPHST
Anolis KKLEKNDIYKQKKISELTVSSGKRQVPANLPREEEDTPTCTQVTDIEDFGVS--RPLQKG
Xenopus KKLEKNDIYKQKRINELFTSEGKRQITAMQMQ----PDSQTPDMEDFGIG--KRLQPA
Danio KKLEKNDIYKQKKISELFTSEGKRQVAQQT-----AAGTTPDMEDLAAP--QRPVQPA
Drosophila KKMLEKNDVLKQRRINEMFTSRPKPKPLAT-----EEDKLADMEDLAGKDGGEAAG
Caenorhabditis NKIRNKFDAHRQPRINQIFAACQKPKSTSQMDNGKRRRTPDDDVA-----SEDAMDSQDD
Saccharomyces RKIATKEDKFKQTSLSLTKFFSKTKNVPTMG-----KIKDIEDLFEPTVEEDNAK

Schizosaccharomyces KRVAVLNSKYQQKKIDSIFSLAPTNPSTI-----NNTKVTDIEDLGSVTHK---DK
.: : . : * : . : *

Homo APVT--VKKRKRVL--WESQEEESQDLTPTVPWQEILGQPALGTSQEEWLWVLRFHKKKWKQ
Mus VPVA--TKRKR-V--WETQKESQDIALTPWQEVLGQPPSLGTTQEEWLWVLRFHKKKWKQ
Anolis VPVS--SKRKRIPAEESQPQSQNLLELTLWSREILGPPPIGTTKEERVAWLCFHKKRWE
Xenopus VPIS--TKRKRVPAEESQGDSDQLALQSWREVLGQPPPLGTTKEERLWVLRFHKKKWE
Danio ILIS--TKRKRISQGEDSQTETQEQELTQSWREILGPPPAMGATREEILVWLRFHKKKWE
Drosophila CPIV--TKRKRISQLEEHDEEA--QPQATTWRQALGAPPPIGETRKTIVEWVRFQKKKWK
Caenorhabditis I I I D--DDKEN---GAKRQKNTK---KVHTTEVVLEK---KTLVEHGFDWMMGFLKKKWR
Saccharomyces IKIARTTKKKAVSKRKRNLQTN-----EEDPLVLPSEIPSMDEDYVGWLNLYQKIKWK
Schizosaccharomyces RIVARVTKRK-----LLQQSGN-----SEAPVSFEVKPVSFMDGYSNWLKYAKKKWK
: . : : : : : * : : * * * .

Homo LQARQRLARRKRQRLESAEGVL-RPGAIRDGPATGLGSFLRRTARSILDLPWQIVQISET
Mus LQAQQLALRRKRQRLESAEDMP-RLGPIREGPSTGLGSFLRRTARSIMDLPWQIIQISET
Anolis LQARQEQERRKRRRLADGEVAR-SGGLIRAGATKGLSNYLLRRTARSILDLPWQIVQIVET
Xenopus LQARQKERQKRRLEDGDVAPGAGGVIRETQAAGLGSFLRRTARSILDMPWQIVQIGES
Danio LQLRQRKERRKRRRLIDGESQPTGGGVIRGGPTTGLGSFLRRTARSILDMPWQIVQIAET
Drosophila WQDQQRQRNRQASKRTRGEDPP---VVRATGSTATLGGFLRRAQRTLLDQPWQIVQLVPV
Caenorhabditis VQRKERKTQLSSKD---S-----DVVEAIVRGAREAEHDKEWHILSVEPT
Saccharomyces IQARDKRRDQLFGNTN-----SSRERSALGSMIRKQAESYANSTWEVLQYKDS
Schizosaccharomyces YQKQVKLRRRHLIGFQS-----R-----QFTNVLQSSAEVMFENLWHILQIRET
* : . : . : * . : .

Homo SQAGLFRWLAVGSDLHCIRLSIPRVFYVN---QRVA-KAEEGASYRKVNRVLRPSN---
Mus RQAGLFRLWAIIGNDLHCIKLSIPRVFYVN---QRVA-KAEDGPAYRKVNRALPSN---
Anolis SQPGLFRLWAVIGSDLHCLKLHVPRIFYVN---QRVP-KPEEGPVYRKVNRILPSN---
Xenopus SQPGLFRLWAVIGNDLHCIKLNI PRVFYVN---QRIP-KPEEGAVYRKEEV-----
Danio SHPLYKLWAVIGSDLHCKMLNI PRVFYVN---QRVP-KQEEAATCKKVNRI LPSG---
Drosophila DDLGHFTVWALIGEELHKIKLTVPRIFYVN---QSAAPPEEGQLWRKVNRLPSR---
Caenorhabditis ADASFFNVWLAVQGMHKVIMKIGRRI IVD---SRAPRGD-----RDTIRRI LPHHK---
Saccharomyces GEPGVLEVFVTINGKVQNI TFH I PKTIYMKFKSQTMPLQKIKNCLIEKSSASLPNNPKTS
Schizosaccharomyces DVPGILHAWVI I RNRLT SIRFIVNRKFFVCFKDETLNVEIEGCLIEKSNAILPHGS---
. : : : : : : . .

Homo --MVYNLYEYSVPEDMYQEHINEINAELSAPDIEGVYETQVPLLFRALVHLGCVCVVNKQ
Mus --IVYNLYEYSVPEDMYQEHINEINTELSVPDIEGVYETQVPLLFRALVQLGCVCVVNKQ
Anolis --LVYNLYEYSVPEEMYQEHLNEINASLSAPDIEGVYETQVPLLLRALI L LGCVCVVNKQ
Xenopus -----NKVK-----TPI-----
Danio --VACFLYQYTPEDMYQEHINEINADLSAPDIEGVYETQVPLLFRALVQLGCMCMVNKH
Drosophila --PVFNLYRYSVPEQLFRDNLGMLADLATPDIEGIYETQMTLEFRALMDMGCICGVQRE
Caenorhabditis --TPGFLYEFRTDENQLTALMDKLYSETCSSTIDGIYSEVPTS FRAVLQ LGSIVRPDHG
Saccharomyces NPAGGQLFKITLPESVFLEEKENCTSI FNDENVLGVFEGTITPHQRAIMDLGASVTFRSK
Schizosaccharomyces --TSDKLFLLEIPEKSYL TEKVSISMI FAHPSVSGIYETRIEPIERLILEMGSRKRFNNS

Homo LVRHL---SGWEAETFALHLEMRSLAQFSYLEP--GSIRHIYLYHHAQ--AHKALFGIF
Mus LTRHL---SGWEAETFALHLEMRSLAQFSYLEP--GSIRHIYLYHHTQ--GHKALFGVF
Anolis LVRHL---TGREGAEFALEHLEMRSLAQCSYLEP--GSIRHIYLYHHTQ--GHKALLGLF
Xenopus -----TDGAI I I LKNT--GSIRHIYLYHNSQ--GHKALFGLF
Danio VVRDL---AGREADTFDLEHLEMRSLAQFSYLEP--GSVRHIYLYHHSQ--GHKALFGLF
Drosophila EARRLAQLATKDLTF SIEQLEQRPQTQVKY LASANNRLRKIYLYQHNTPTAKKEIWSLI
Caenorhabditis ISL-----GG---HQLTLENL--KPMEKAPYL-PLDQKIRTI FLYKFSQD--SRHVYSLI
Saccharomyces AMGAL---GKGIQQGFEMKDL SMA--ENERYLSG--FSMDIGYLLHFPT--SIGYEFFSL
Schizosaccharomyces VPGAL---GKGFEFGFESKMF TDPDNDVSYLDG--VEMNYLYAFHFSI--SNRFVFSLF
: : : :

Homo IPSQRRASVFLVDT----VRSNQMPSLGALYSAEHGLLLEKVG---PELLPPPKHTEFEVR
Mus IPSQRRASVFLVDT----VRSNQMPGLSALYSSEHSLLLDKVD---PKLLPPPKHTEFEVR
Anolis IPSQRRASVFLVDT----VRSNQMPSLANMYTAERNAEAEKVD---PDLLPPDKHTEFEVQ
Xenopus IPSQRKASVFLVDT----VRSNQMPNLSNMYTAETAMQERVD---PELLPPEKHIFEVQ
Danio IPSQRKASIFVLDT----VRSNQMPNLSNLYGAERTALLEKTT---EELLPPEKHIFEVQ
Drosophila LMPSKKAFVFDLDT----VRANQMPNMRQLYTAERLALLKNTLAEEQDKI PVEDYTFEVL
Caenorhabditis DSSG-SAAIFYIVN----TGDVQMPNMDSLYTSAYTKMMSTERGQ--LCHTSESMPTTVK
Saccharomyces FKSXGDTITILVLKPSNQAQEI NASSLQGIYKQMF EKKKGIETYSYLVDIKEDINFEFV
Schizosaccharomyces MPHLLKVEAIIYDK--LPGSDMSFPSISKIYEELRSKFDNLIKES--IEYPTLSCNVI

: . . : :*

Homo AETDLKTCRAIQRFLLAYKEERRGPTLIAVQSSWELKRLASEIPVLEEFPLVPICVADK
Mus AETNLKTCRAIQRFLLAYKEERRGPTLIAVQSSWELCLRTSEIPVLEEFPLVPIRVADK
Anolis VETDRRAIYRAIHRLLLAYKDERRGPTLIAVQSNWDLKRLASGIPIFEEFPLVPIQVADN
Xenopus AETDLKTVYRGIQRLLLLSYKDERRGPTLIAVQSNWNLRLRAAGMPVLEEFPLVPVIRITDD
Danio AENDIKAIYRALQRILLNYKEERRGPTLIAVQSNWELRRLAAGMTVLEEFVVPVHVIDE
Drosophila IEVDVKQIYRHIQRALTTYKQEHQGPITLCLQTALSARKLSLAMPILLEFPQAEIHISDD
Caenorhabditis RFSSNTECERQLGRALRVYREVSSKTAIVLLSDDTDFRRLARKLPNLGLFPNVQLHITEP
Saccharomyces YFTDISKLYRRLSQETTKLKEERGLQFLLLLQSPFIT-KLLGTIRLLNQMPIVKLSL--N
Schizosaccharomyces FSGNERKAYKLIDEKLLQYFSTKTKNLLIIESSLPH-ILKANVKQIEELPYIMIPR--L
. : : . . : : : * : : :*

Homo IN-YGVLWDQRHGARRMIRHYLNLDTCLSQAFEMSRYFHPIGNLPEDISTFGSDLFFAR
Mus IS-YAVLDWQRHGARRMIRHYLNLDLCLSQAFEMSRYFHPIVGNLPEDISTFGSDLFFAR
Anolis VN-YGILDWQRHAARHLIRRYLNLDTCLSQVFEMSRYYHPIGNLPEDISTFGTDLFFSR
Xenopus IS-YGVLWDQRHGARRMIRHYLNLDSCLSQAFEMARYYHLPVGNLPDDVSTFGSDLFFSR
Danio IS-YNVLDWQRHGARRMIKHYLNLDSCLSQAFDMARYYHLPVGNLPQDISIFGSDLFLAR
Drosophila ASLLSGLDWQRQGSRAVIRHFLNLDLMLDQCRYFHVPIGNMPPDVTVLFGADLFFAR
Caenorhabditis SLLNQIDWQKVVARRVLQHYFNSFFFADYLEWARYLRVPIGNLPAHALFGLDLFFAR
Saccharomyces EVLLPQLNWQPTLLKLVNHVLSGWSHSLIKLSQYSNIPICNLRDSDMYIIDVLYAR
Schizosaccharomyces ESNIQSLSWKQHIATKMIQHFLAIGSWLFHRIQLSRFSDIPLCNFESDDIQYSIDVVYSR
:.*: :.: : : : : : * : : :*

Homo HLQRHNHLLWLSP-TARPD LGGKEADDNCLVMEFDD---QATVEINSSGCYSTVCVELDL
Mus HLQHNHLLWLSP-TSRPD LGGKEADDNRLVMEFDD---RATVEINSSGCYSTVCVELDI
Anolis HLGRHNHLLWLSP-AVRPD LGGKEADDSRLVMEFDE---KISVEINHPGCYSTVCVELDI
Xenopus HLRHNHLLWLSP-SARPD LGGKEADDNRLVMEFDE---KASVEINHPGCYSTVCVELDI
Danio HLRKHNHLLWLSP-TARPD LGGKEADDSRLVMEFDE---RGSMEINSSGCYSTVCVELDI
Drosophila LLQRHNHLLWLSP-STRPD LGGKEADDSRLVMEFDE---SISVVQNKAGFYPDVCVELAL
Caenorhabditis NLQKSGHALWATR-ASRPD LGGKEADDSRLVMEFDE---RGSMEINSSGCYSTVCVELDI
Saccharomyces KLKKNENIVLWVNEKAPLPDHGGIQNDFFD-LNTSWIM-NDSEFPKINNSGVYDNNVLDVGV
Schizosaccharomyces KLKEHNIILWVWNK-GTPD LGGKEADDSRLVMEFDE---RGSMEINSSGCYSTVCVELDI
* . ** . ** * : * * * . : : : :

Homo QNLAVNTILQSHHVNDMEGADSMGISFDVVIQASLEDMITGGQAASAPASYDETALCSNT
Mus QNLAVNTILQSHHVNDMEGAGSMGISFDVVIQASLEDMVTGNQAASALANYDETALCSST
Anolis QSLAVNTILQSHHVNDMEGAASTCVSFDVVIQASLEDMVTGNQAASVPASYDEAALCSNT
Xenopus QSLAVNSILQSHHINDLEGAGSMGISFDVVIQASLEDMVTGNQA--SSVASYDETALCSNT
Danio QSLAVNTILQSQHVNDMEGGASIGVSFDVVIQHASLEEMMSGNQGANAASVYDETALCSNT
Drosophila DSLAVSALLQSTRIQEMEGAS-SAITFDVMPQVSLEEMIGTVP-AATLPSYDETALCSAA
Caenorhabditis SAVAVTALVQRSRVLEAEGADD-VVTFDSMNTIAQQSVTGG--AVNSIACYDEGAAVDAT
Saccharomyces DNLTVNTILTSALINDAEGSDLVNN----NMGIDDK-----DAVINSPEFVHDAFSDNA
Schizosaccharomyces SNLALCSILNSALINDIEGIGDMAALNDNYMTAIN-----D-DLEEKLG IHDNIGLTHS
. : : : : : : : * : : * : : : * : : : *

Homo FRILKSMVVGWVKEITQYHNIYADNQVMHFYRWLRS PSSLHDPALHRTLHNMMKKLFLQ
Mus FRILKSMVVGWVKEITQYHNIYADNQVMHFYRWLRS PSSLHDPALHRTLHNMMKKLFLQ
Anolis FRILKSMVVS WVKEITQYHNVYADNQVIHFYRWLRS PSSLHDPALHRLHMLHNMMKKLFLQ
Xenopus FRILKSMVVGWVKEITQYHNVYADNQVMHFYRWLRS PSSLHDPALHRTLHNMMKKLFLQ
Danio FRILKSMVVGWVREITQYHNVYADNQVMHFYRWLRS PSSLHDPALHRTLHNMMKKI FLQ
Drosophila FRVMSMVNGWLREVSINRNI FSDQIVHFYRWVRS SNALLYDPALRRSLNMLRKMFLR
Caenorhabditis IKVLKQMLTECVRHIAHQGNARADEVMTVSRWLNTRS ALLFDAALTRSVSVLESKLVLL
Saccharomyces LNVLRGMLKEWWDEAL-KENSTADLLVNSLASWVQNPNAKLF DGLLRVHVNHLTKKALLQ
Schizosaccharomyces LPVLKALVKTTWWNEAA-SGNNLADLI IQHLARWISSKSYLSPLSSHV EIVMRKTFLO
: : : : : . * : * : . * : : : * . * : : * . *

Homo LIAEFKRLGSSVIYANFNRIILCTKKRRVEDAIAYVEYITSSIHSKETFHSLTISFSRCW
Mus LIAEFKRLGSSVYANFNRIILCTKKRRIEDALAYVEYITNSIHSKEIFHSLTISFSRCW
Anolis LVAEFKRLGSLVYANFNRIIVLCTKKRHIEDALAYVEYITNSIHSKEIFHSLTISFSRCW
Xenopus LVAEFKRLGSSVYANFNRIILCTKKRRVEDAVAYVEYITQSIHSREIFHSLTISVSRCW
Danio LVAEFKRLGSSVIYANFNRIILCTKKRRIDDAIAYVEYITNSIHTKEIFHSLTISFSRCW
Drosophila LIAEFKRLGATIIYADFNRIILSSGKKTVDALGYVDYIVQSLRNKEMFHSIQLSFEQCW
Caenorhabditis LCAECERIGAKVIHATAQKLVLNLTGKSTSEEAKGFAEMLIQSLSTNVVFAALHITPVKFF
Saccharomyces LVNEFSALGSTIVYADRNQILIKTNKYSPENCYAYSQYMMKAVRTNPMFSYLDLNIKRYW
Schizosaccharomyces LLSEIKRLGAHIIHASANKILIKTSKLVQNAVYTSNYLLKSIKTLPLFHFLLDNVTEYW
: * . : * : : : : : * : : : : : * : : . : :

Homo EFLLWMDPSNYGGIKGKVSSRIHCGLQDSQK---AGGAE---DEQENEDDEEERD--GEE
Mus EFLLWMDPSNYGGIKGKVPSSIHCGQVKEQD---SQARE---ETDEEE--EDKEK--DEE
Anolis EFLLWMDPANYGGIKGKVP SRVHC GEGTNSK---PTADA---DGSEDEEEEEE-----DK
Xenopus EFLLWMDPANYGGVKGKLPSSVHYGEKKK---ALDGEDG---DGSEDEEEEDDQPV--NSD
Danio EFLLWMDPANCGGVKGKLPSSLLYGEHDSEKKKKDGAE---DGSDDEDEHEENQ--EED
Drosophila NFMLWMDQANFSGIRGKLPKGI DETVSSIVSTT MIRDSE RNQDDDEDEEEDSEN RDPVES
Caenorhabditis DAMLWMDAHNHTGIRISEKTE-----SS-----PDVIA-----
Saccharomyces DLLIWMDKFNFSGLACIEIE-----
Schizosaccharomyces DYLLWMDSVNYGGK MVAANF-----
: : : * * *

Homo EEEAEESNVEDLLENWNI LQFLPQAASCQNYFLMIVSAYIVAVYH CMKDGLRRSAPGST
Mus EEGMGESSEVEDLLENWNI LQFLPQAASCQSYFLMIVSAYIVAVYQSMKEELRHSAPGST
Anolis EEGDGAEGVEDLLENTWNIVQYLPQAASCQTYFLMIVSAYIVAVYHSLKEERQ RNSPGST
Xenopus -DEQSEDQVEESLENNWNI VQYLPQAASCQSYFLMIVSAYIVAVYHSMREEMRRNAPGNT
Danio EEQDREEDVEDLIESNWNIMQYLPQTASCQKYFLMI ISAYIAAVYHSMRQELRRNAPGAT
Drosophila NEAEQDQDELSLELNWTIGEHL PDENECREKFESLLT LFMQSLAEKKTTE-----
Caenorhabditis -DEEES SCTEFETTAIWKIAEEMPT EANIQEEFLQ MIGAYILEFLETNRKMHF DSESG--
Saccharomyces ----EKENQDYTAVSQWQLKKFLSPI--YQPEFEDWMMI I LDSMLKTKQSYLKLNSGTQR
Schizosaccharomyces ----SATNEEPQTVVSWHIKSHL PPI--IQPEFQSWIVEFIEEVYKQKLEKSNTKVG---
: * : : * : : .

Homo PVRRRGASQLSQEAEGAVGALPGMITFSQDYVANELTQSFFTITQKI QKKVTG--SRNST
Mus PVKRRGASQFSQESEGATGSLPGMITFSQDYVANELTQSFFTITQKI QKKVTG--SRNTT
Anolis PVKRRASSQVSQEPAEERGAMPGTIAFSQDYVANELTQSFFTITQKI QKKVSG--SRRTT
Xenopus PIKRRQNSQVSQEANADGGAMPGLITFSQDYVANELTQNFFTITQKI QKKVTG--TRHAT
Danio PIKRRGATQASQQPSGDVSALPGLITFSQEYVSSELTQNFFTITQKI QKKVSS--TRSVT
Drosophila -----Q-----AI-----KDISHCAFDFILK LHKNY-----
Caenorhabditis -----ATFRSDCISQKISHRLYRIVNKMVHN NAD-----
Saccharomyces PTQIV---NVKKQDKEDSVEN-SLNGFS-----HLFSKPLMKRVK KLFKNQQEFILDPQY
Schizosaccharomyces ---FV---RVKNNNADEDSEIVGSGILK-----SKLIHPLKRRVAQVRRCFQELQLDENT
: : : : :

Homo ELSEMFPVLP GSHLLLNPALEFIKYVCKVLSLDTNITNQVNKLN RDLRLVDVGEFSEE
Mus EPSEMFPVLP GSHLLLNPALEFIKYVCKVLSLDTNITNQVNKLN RDLRLVDVGEFSEE
Anolis KPSAMFPSP LPGSYLVFNPALEFVKSVCEVLSLDANVTN QVSKLKRDLRLVDVREFSEE
Xenopus EPSEMFPILPGSHLPLNNPALEFIKYVCEVLALDSNITNQVNKLN RDLRLVDVGEFSED
Danio LPSEMFPVLP GSHLPLNNPALEFIKYVCQVLSLDANIVNQVNKLN RDLRLVDVGEFSED
Drosophila -----GKGKPSG LELIRTLIKALSVDKTLAEQINELRRNMLRLV GIGEFSDL
Caenorhabditis -----IAHCSVYLANALCRALSCDQTSQLAVEGIRDNAKRL LHNSVVEAD
Saccharomyces EADYVIPVLP GSHLVNKNPLLELVKSLCHVM LLSKSTILEIRTLR KELLKIFELREFAKV
Schizosaccharomyces REDLKF PKLP GSFNLNYTDGALELVK SICAVFELSHDLNLEVRFLKKSLLSLLQIQEFSTQ
: : . : . : . : .

Homo AQFRDPCRSYVLP EVICRSCNFCRDLDLCKDSSFS EDGA-----VLPQWLC SINCQAPYD
Mus AQFRDPCHSYVLP EVICRSCNFCRDLDLCKDSSFSQDGA-----ILPQWLC SINCQAPYD
Anolis AQFQDPCCSYVLP EMICKSCNFCRDLDLCKDPVFSQDGS-----VLPGWAC PSCQAPYD
Xenopus AQFRDPCRSYILPEI ICRCNFCRDLDLCKDPAINQDGS-----VLPQWVC TNCQAEYD
Danio AQFRDPCKSYVLP EVICHCNFCRDLDLCKDPAVSQDGS-----VLPQWFCSNCQ AQYD
Drosophila AEWEDPCDSH I NEVICACNHC RDLDLCKDKHRAMK-D-----GVPVWLCAQCYVAYD
Caenorhabditis MT-PLRSTTLFVSNVFCNSCSQASNVFLS-----S-----TDEILTCATCQSKLN
Saccharomyces AEFKDP SLSLV PDLFCEYCF FIDIDFCKAAP-----ESIFSCVRCHKAFN
Schizosaccharomyces AVFRYPSRRLSLDQIPCKQCGVHQDFDLCLHEHLWPTRDDMGTLVFS DGSWSSCNLVYD
: : . * . * : : . * * :

Homo SSAIEMTLVEVLQKKLMAFTLQDLVCLKCRGVKETSMPVYCSCAGDFALT IHTQVFMEQI
Mus SSAI ESALVEALQ RKLMAFTLQDLVCLKCRGMKETHMPVYCSCAGDFTLTIRTEVFMEQI
Anolis SDAIELALVEALQKKLMAFSLQDLVCLKCKG I KDSHMLHCSAGDFD LLLPTKTFLDQL
Xenopus SDGIEMALVEALQKKMMAFTLQDLVCMKCKGVKEANMPVYCNCAGDFTLTISVKV LLEQI
Danio TESIEMALVEALQKKLMSYTLQDVECAKCKGVKEANMPLYCSCAGDFNLFTT KSFSEQV
Drosophila NEEIEMRMLDALQRKMLSYV LQDLRCSRCEIKRENLA EFCTCAGNFVPLISGKDIQTL L
Caenorhabditis SDVIDMMRICDRLNQLLTAYQIQDHQCTKCKSVRHD TLSMYHCSCS QFIPQITPAQLKHEA
Saccharomyces QVLLQEH LIQKLRSDIESYLIQDLRCSRCHKV KRDYMSAHCPCAGAWFETIPRESIVQKL
Schizosaccharomyces RWVFEETLV DNLYHQLTLYQLQDLICSKCKTVKQWSLKERCSGSEWV LQLSPTKFREML
: : : * : : : * * * : * : : * * . : : :

Homo	GIFRNIAQHYGMSYLLETLEWLLQKNPQLGH
Mus	RIFQNIAKYYSMSYLQETIEWLLQTSFVSNC
Anolis	KVFQSI AQHYNMAYLLDTVTWLLGKSPRLPS
Xenopus	NVFRNIAQHYNMAHLLLETIVWIMQQNSHLIQ
Danio	EVFRNIAAHYNMSFLLETIDWII SMNT-----
Drosophila	GTFNKVAANHMKQLLQQT VHQAL TTPR-----
Caenorhabditis	STVETVSIVRNFALSSELATWVLKML-----
Saccharomyces	NVFKQVAKYYGFDILLSCIADLTI-----
Schizosaccharomyces	NVYQSVADFYEF SILQNSVQSILSVLN-----
	. :: : .

Supplementary Figure 2b. Clustal omega multi-sequence alignments of POLE across species

Residue 478 in humans is highlighted.

CLUSTAL O(1.1.0) multiple sequence alignment

```
Homo      -----MDGKRRP-GPGPG
Mus       -----MDCKRRQ-GPGPG
Anolis    -----MG
Xenopus   -----
Danio     -----MDFKRRNGGPAMG
Drosophila -----MDGKRKFNGTSNG
Caenorhabditis -----MTSKR-PGGSSFQ
Saccharomyces MSEKRS LPMVDVKIDDEDTPQLEKKIKRQSIDHGVGSEPVSTIEIIPSDSFRKYNSQGFK
Schizosaccharomyces MTRSSNEG--VVLNKENYPPFR----RNGSIHG-----EITDVKRRRLSERNGYG
```

```
Homo      VPPKRARGGLWDDDDAPRPSQFEEDLALMEEMEAEHRLQEQEEEEELQSVLEGVADGQVPP
Mus       VPPKRARGHLWDED-EPSPSQFEANLALLEEIEAENRLQEAE-EELQLPPEGTVGGQFST
Anolis    EPTLSQRKKS RGDWEEDAPSQFEEELAF LDEVESEMALEA---KETQLVPDDIPLGNLFS
Xenopus   -----
Danio     GASQAKKKGK P STEWE-DSPSQFEEELALFDEMMDM--EVES---GEGQAGHDV I PVGDLFS
Drosophila HAKKPRN-----PDDDEEMGF EAELAAFENSEDMQ-----TL-----LMGDGPEN
Caenorhabditis --PEVKRKRRESDEFEQCYVSRFENELPS-----VP
Saccharomyces AK-----DTDLMG TQLESTFQEELS QMEHDMAD-----QEE-
Schizosaccharomyces DK-----KG-SSSKEKTSSFEDELAEYASQLDQ-----DEIK
```

```
Homo      SAIDPRWLRPTPP-ALDPQTEPLIFQQLEIDHYVGPAQVPVGGPPPSRGSVPVLRAFGVT
Mus       ADIDPRWRRPTLR-ALDPSTEPLIFQQLEIDHYVGSAPPLPEGPLPSRNSVPILRAFGVT
Anolis    SVRNPKWQRPPPP-EINPKEDALCFQQVELDYVVGTHI--PGFPGSTQGPVPLRMFGVT
Xenopus   --IPPKWLRPPVS-LTDPKEQHLCFQQVELDHYVGVSHV--SGMLGATKGPVPIIRMFGIT
Danio     TDLNPRWKRPHAS-PLHPKSDTLIFQQIDL DYLGAAV--AGMPGQVQKVP I VRMFGVT
Drosophila QTTSERWSRPPPP-ELDP SKHNLEFQQLDVENYLGQ--PLPGMPGAQIGPVPVVRMFGVT
Caenorhabditis TIDKTGWARPAVD-KDLGISKSIACQILEVETYHEDGS-----ATSYDRTNVKLYGVT
Saccharomyces -HDLSSFERKKLPTDFDPSLYDISFQQIDAEQSVLNGI-----KIDENTSTVVRFFGVT
Schizosaccharomyces SSKDQQWQRPALP-AINPEKDDIYFQQIDSEEF T-----EGSVPSIRLFGVT
          : *           : * :: :           :: : * : *
```

```
Homo      DEGFSVCCHIHGFAPYFYTPAPPFGFPEHMGDL---QRELNLAI SRDSRGGREL-----
Mus       DEGFSVCCHI QGFAPYFYTPAPPFGFAEHLSEL---QRELNA AISRDRGGKEL-----
Anolis    EAGNSLCCHIHGFAPYFYVPAQTGFKADHLAEF---QRELNA AVLDRMRSNKDN-----
Xenopus   EEGNSVCCHIHGFAPYFYVPCHTGFKQEDLSDF---KKELN TAVIKDMRSNKDG-----
Danio     DNGNSVCCHIHGFAPYFYVPAPNGFTNAHLAGF---QRELNSVVLKDMRSNKDN-----
Drosophila MEGNSVCCHVHGF C PYFIEAPSQFEEHHCEKL---QKALDQKVIADIRNNKDN-----
Caenorhabditis KSGNSICVIVTDYFPHFYFQAPQGFGEVHIGTA---QSAICNMVAAA KRRGGSGQAQLPG
Saccharomyces SEGHSVLCNVTGFKNYLYVPAPNSSDANDQEQINKFVHYLNETF-----
Schizosaccharomyces DNGNSILVHVVGFLPYFYVKAPVGRPEM---LERFTQDL DATC-----
          * *: : : : * .           :
```

```
Homo      -TGPAVLAVELCSRESMFGYHGHG-PSPFLRITVALPRLVAPARRLLEQGIRVA-G-LGT
Mus       -SGPAVLAI ELCRESMFGYHGHG-PSPFLRITLALPRLMAPARRLLEQGV RVP-G-LGT
Anolis    -LSQVVLAVEICNKQNM YGYHGEN-FIPFLKITMAMPRLIAPAKR LLEQGLRCG-T-LGV
Xenopus   -ISQAVLAVDVCHKENMYGYHGKR-IMPFMKITMALPRLIAPAKR LLEQGLRFG-R-HPI
Danio     -IAVTVLAVDITKKESMYNYHGNK-PHDFLRITMAMPRLVAPAKR LLEQGFKFA-N-FAT
Drosophila -VQEAVLMVELVEKLN IHGYNGDK-KQRYIKISVTLPRFVAAA S RLLKKEVIMS-E-IDF
Caenorhabditis KVVDNLVHVEIVHGENLYYFRGADTKVPFVKVSGSTEALH-KARME LKNGVNL MGK-GPV
Saccharomyces -D-HAIDSIEVVSKQSIWGYSGDT-KLPFWKIYVTYPHMVNKLRTAFERGHLSFN S WFSN
Schizosaccharomyces -NGGVIDHCI IEMKENLYGFQGN E-KSPFIKIFTTNPRILSRARNVFERGEFNFEELFPV
          : : : * : : : : :
```

```
Homo      -PSFAPYEANVD FEIRFMVDTDIVGCN WLELPAGKYALRLKE-----KATQCQL
Mus       -PSFAPYEANVD FEIRFMVDADIVGCN WLELPAGKYVRRAEK-----KATLCQL
Anolis    -HNYQAFEANIDFEIRFMVDRDVGCN WIELPAGKYRLRQE QSAGESSKENPPKVS L CQL
Xenopus   -HCYQAYEANIDFEIRFMVDNDIVGCN WIELPAGKYRVRKESQDEEPSKDNPSKVS L AQI
Danio     -QSYQAFEANIDFEIRFMVDSNVVGC CWIELPKGKYQLREERREGQ TDSKYPGKVS L CQY
Drosophila -QDCRAFENNIDFDIRFMVDTDVGCN WIELPMGHWRIRNSHS-----KPLPESRCQI
```

Caenorhabditis	-NVGNLYESNINVIVMFLAKTNIVGCGWIEIPAGKCRILSNSE-----KSSRCQI
Saccharomyces	GTT---TYDNIAYTLRLMVDGIVGMSWITLPGKYSMIEPN-----RVSSCQL
Schizosaccharomyces	GVGVTTTFESNTQYLLRFMIDCDVVMNWIHLPAASKYQFR-YQN-----RVSNCQI
	* : : . : ** * : * . :
Homo	EADVLSVDVVSHPPEGPWQRIAPLRVLSFDIECAGRKGIFPEPERDPVIQICSLGLRWGE
Mus	EVDVLSVDVISHPPEGEWQRIAPLRVLSFDIECAGRKGIFPEPERDPVIQICSLGLRWGE
Anolis	EADVAWMDFISYPPEGEWQRIAPLRVLSFDIECAGRKGIFPEPEKDPVIQIANMVQRQGE
Xenopus	EVDISWADLISHPAEGEWQKIAPQRVLSFDIECAGRKGIVFPEPEKDPVIQIANMVLRQGE
Danio	EVDVAWDSLISHPAEGEWQRIAPLRVLSFDIECAGRKGIVFPEPEVDPVIQIASMVQRQGE
Drosophila	EVDVAFDRFISHEPEGEWSKVAPFRILSFDIECAGRKGIFPEPEKDPVIQIANMVIRQGE
Caenorhabditis	EVTVPVKNLIVHESDGEWAGIAPIRTLSDIECIGRRGVFPEAKIDPVIQIANLVKIEGE
Saccharomyces	EVSINYRNLIHAPAEGDWSHTAPLRIMSFIECAGRIGVFPEPEYDPVIQIANVVSIAAGA
Schizosaccharomyces	EAWINYKDLISLPAEQWSKMAPLRIMSFIECAGRKGIVFPDPSIDPVIQIASIVTQYGD
	* . : . : : * * * * * : * : * * * : * : * * * : * : * * * : * : *
Homo	PEPFLRLALTLRPCAPILGAKVQSYEKEEDLLQAWSTFIRIMDPDVTGYNIQNFDPYLL
Mus	PEPFLRLALTLRPCAPILGAKVQSYEREEDLLQAWADFI LAMPDVTGYNIQNFDPYLL
Anolis	KDPFVRNVFTLQSCAPIVGSQVLCFQKEGELLKAWAEFIRTVDPDIITGYNIQNFDPYLL
Xenopus	KDPFIRNVFTLGTCSAPIVGSQVLCFEREDALLKAWAEFVRIIDPDIITGYNIQNFDPYLL
Danio	KEPFIRTVFTLQSCASIVGSQVLCFTQEKMLQSWAEFMRTVDPDIITGYNIQNFDPYLL
Drosophila	REPFIRNVFTLNECAPIIGSQVLCCHDKETQMLDKWSAFVREVDPDILTGYNINNFDFPYLL
Caenorhabditis	AEPFVRNCFVLGTCAVVGSNIIQCVNEKVLLEKWAEFVREVDPDIIITGYNINNFDFPYLL
Saccharomyces	KKPFIRNVFTLNTCSPIITGSMIFSHATEEEMLSNWRNFIKVDPDVIIGYNTTNFDIPYLL
Schizosaccharomyces	STPFVRNVFCVDTCSQIVGTQVYEFQNAEMLSWSKVFVRDVPDVLIGYNICNFDPYLL
	** : * : : * : * : : : : * . * * : : * * * : * * * * * :
Homo	ISRAQTLKVQTFPFLGRVAGLCSNIRDSSFQSKQTGRDRTKVSMVGRVQMDMLQVLLRE
Mus	ISRAQALKVDRFPFLGRVTGLRSNIRDSSFQSRQVGRDRSKVISMVGRVQMDMLQVLLRE
Anolis	INRAQTLKVSTFPFLGRIPARKSVIRDSSFQSKQMGRENKVINTEGRAQFDLLQVLLRD
Xenopus	INRAQTLKVSTFPFLGRIRSLKSVIRDSSFQSKQMGRENKVINTEGRVQFDLLQVLLRD
Danio	LNRAATLKVNMFPYLGRVWGCKSVLKDSSFQSKQMGRENKVINTEGRVQFDLLQVLLRD
Drosophila	LNRAAHLKVRNFEYLGRINKRISVIKEQMLQSKQMGRENQYVNFEGRVPFDLLFVLLRD
Caenorhabditis	LDRAKVLSLPQVSHLGRQKEKGSVVRDAAISSKQMGSRVNKSIDIHGRIIFDVLQVLLRD
Saccharomyces	LNRAKALKVNDFPYFGRKTKVQEIKEKSVFSSKAYGTRETKNVNIDGRQLDQLQFIQRE
Schizosaccharomyces	LDRAKSLRIHNFPLLGRIHNFVSAKETTFSSKAYGTRESKTTSPGRQLDMLQVMQRD
	: . * * * : . : * * . : : : * : * * . : . * * : * * * : * * :
Homo	YKLSYTLNAVSYFHFLGEQKEDVQHSIITDLQNGNDQTRRRRLAVYCLKDAYLPLRLLERL
Mus	HKLSYTLNAVSYFHFLGEQKEDVQHSIITDLQNGNEQTRRRRLAVYCLKDAFLPLRLLERL
Anolis	YKLSYTLNAVSYHFLOEQKEDVQHSIITDLQNGTEQSRRRRLAVYCLKDAYLPLRLLLEKL
Xenopus	YKLSYTLNAVSYHFLOEQKEDVQHSIITDLQNGNEQTRRRRLAVYCLKDAYLPLRLLLEKL
Danio	YKLSYTLNAVSYHFLOEQKEDVQHSIITDLQNGNEQTRRRRLAVYCLKDAYLPLRLLQKL
Drosophila	YKLSYTLNAVSYHFLOEQKEDVHHSIITDLQNGDEQTRRRRLAMYCLKDAYLPLRLLLEKL
Caenorhabditis	YKLSYTLNAVSYQFLSEQKEDVEHNIIPDLQRGDEQTRRRRLAVYCLKDAYLPLRLLDKL
Saccharomyces	YKLSYTLNAVSAHFLGEQKEDVHYSIISDLQNGDSETRRRRLAVYCLKDAYLPLRLLMEKL
Schizosaccharomyces	EKLSYSLNAVCSQFLGEQKEDVHYSIITDLQNGTADSRRRLAVYCLKDAYLPLRLLMEKL
	. * * * * : * * * . : * * * * * . : . * * * * * : * * * * * : * * * * * :
Homo	MVLVNAVEMARVTGVPLSYLLSRGQVKVVSQLLRQAMHEGLLMPVVKSE-GG-EDYTGA
Mus	MVLVNNVEMARVTGVPLGYLLTRGQVKVVSQLLRQAMRQGLLMPVVKTE-GS-EDYTGA
Anolis	MCVINMEMARVTGVPLSYLLARGQQIKVVSQLLRQAVKQNLVMPVVKSE-GG-EDYAGA
Xenopus	MCVINMEMARVTGVPLSYLLSRGQIKVVSQLLRQAMKQDLVMPVVRSE-GG-EDYTGA
Danio	MCVINMEMARVTGVPLTYLLSRGQIKVVSQLLRQAMKQDLVMPVVKTE-GG-EDYTGA
Drosophila	MAIVNMEMARVTGVPLESLLTRGQIKVLSQLLRQAKTKGFIMPYSYSQ-GSDEQYEGA
Caenorhabditis	MSIINYIEMARVTGVPMNLLTKGQIKILSMMMLRCKQNNFFLPVIEANS GDGEGYEGA
Saccharomyces	MALVNYTEMARVTGVPFYSYLLARGQQIKVVSQVFRKCLEIDTVIPNMQSQ-ASDDQYEGA
Schizosaccharomyces	MCFVNYTEMARVTGVPFNLLARGQQIKVISQVFRKALQHDLVVPNIRVN-GTDEQYEGA
	* . : * * * * * * * : * * : * * * * : * * : * * . : * * : . : * * *
Homo	TVIEPLKGYDVPDIATLDFSSLYPSIMMAHNLCTTLLRPGTAQKL----GLTEDQFIRT
Mus	TVIEPLKGYDVPDIATLDFSSLYPSIMMAHNLCTTLLRPGAAQKL----GLKPDEFIKT
Anolis	TVIEPLKGYDVPDIATLDFSSLYPSIMMAHNLCTTLLQGGAVERH----GLSADKFIRT
Xenopus	TVIEPLKGYDVPDIATLDFSSLYPSIMMAHNLCTTLLQSGSVEKY----GLNPDEFIKT
Danio	TVIEPEKGYDVPDIATLDFSSLYPSIMMAHNLCTTLLQKSQIEKL----GLGPDDFIKT
Drosophila	TVIEPKRGYADPISTLDFASLYPSIMMAHNLCTTLLVGGTREKLRQQENLQDDQVERT
Caenorhabditis	TVIDPIRGFYNEPIATLDFASLYPSIMIAHNLCTTLLKSP-----QGVENEDYIRT

Saccharomyces TVIEPIRGYYDVPIATLDFNSLYPSIMMAHNLCTTLCNKATVERLNL---KIDEDYVIT
Schizosaccharomyces TVIEPIKGYDTPATLDFSSLYPSIMQAHNLCTTLLDSNTAELLKL---KQDVDYSVT
***:* :*: * **:* ** * ** * ** * ** * ** * ** *

Homo PTGDEFVKTSVRKGLLPQIENLLSARKRAKAEELAKETDPLRRQVLDGRQLALKVSANSV
Mus PTGDEFVKS SVRKGLLPQIENLLSARKRAKAEELAQETDPLRRQVLDGRQLALKVSANSV
Anolis PTGDHFVKASVRKGLLPEILENLLAARKRAKTELKQETDPFKRQVLDGRQLALKVSANSV
Xenopus PTGDFVKASVRKGLLPEILENLLCARKRAKTELKKETDPFKQVLDGRQLALKVSANSV
Danio PTGDLFVKSSVRKGLLPEILENLLSARKRAKAEELKKETDPFKKQVLDGRQLALKISANSV
Drosophila PANNYFVKSEVRGGLLPEILESLAARKRAKNDLKVETDPFKRQVLDGRQLALKISANSV
Caenorhabditis PSGQYFATKSKRRGLLPEILEDLAARKRAKNDMKNEKDEFKRMVNGRQLALKISANSV
Saccharomyces PNGDYFVTKRRRGLLPIILDELISARKRAKDLRDEKDPFKRDLVNGRQLALKISANSV
Schizosaccharomyces PNGDYFVKPHVRKGLLPIILADLLNARKKAKADLKKETDPFKKAVLDGRQLALKVSANSV
* : *.. **:* ** * .: : **:* ** : : *.* : : * :*****:*****

Homo YGFTGAQVGKLPCLLEISQSVTGFRQMIKTKQLVESKYT---VENGYSTSAKVYGD TD
Mus YGFTGAQVGKLPCLLEISQSVTGFRQMIKTKQLVESKYT---VENGYDANAKVYGD TD
Anolis YGFTGAQVGKLPCLLEISQSVTGFRQMIKTKQLVESKYT---IANGYSADAKVYGD TD
Xenopus YGFTGAQVGKLPCLLEISQSVTGFRQMIKTKQLVESKYT---LDNGYKADAKVIYGD TD
Danio YGFTGAQVGKLPCLLEISQSVTGFRQMIKTKQLVESRYT---LDNGYQADAKVIYGD TD
Drosophila YGFTGAQVGKLPCLLEISQSVTAYGRMIEMTKNEVESHYT---QANGYENNAVVIYGD TD
Caenorhabditis YGFTGATVGKLPCLLEISQSVTAFGRKIMDMTKLEVERIYKKGALDGKCPADAKVIYGD TD
Saccharomyces YGFTGATVGKLPCLLAISSSVTAYGRMILKTKTAVQEKYC---IKNGYKHDVAVVYGD TD
Schizosaccharomyces YGFTGATNGRLPCLLAISSSVTSYGRQMIKTKDVVEKRYR---IENGYSHD VAVVIYGD TD
***** *:* ** * ** * .:* ** * ** * : * * . * * :*****:*****

Homo SVMCRFGVSSVAEAMALGREAADWVSGHFSPPIRLEFEKVIYFPYLLISKKRYAGLLFSSR
Mus SVMCRFGVSSVAEAMSLGREANWVSSHFPPIRLEFEKVIYFPYLLISKKRYAGLLFSSR
Anolis SVMCR LGVPSVAEAMEIGREAAWVSSHFIPIKLEFEKVIYFPYLLINKKRYAGLYFSSN
Xenopus SVMCKLGVTVAEAMELGREAAEWVSSHFTPIKLEFEKVIYFPYLLINKKRYAGLYFSSS
Danio SVMVKLG VATVQ EAMNQGKEAAEWVSSHFPPIKLEFEKVIYFPYLLINKKRYAGLYFSSS
Drosophila SVMVNFVGT LERSMELGREAAELVSSKFVHP IKLEFEKVIYFPYLLINKKRYAGLYFT-R
Caenorhabditis SVMVKFGVETVAQAMEIGLDAAKEVSKIFTPIKLEFEKVIYFPYLLINKKRYAGLYFT-K
Saccharomyces SVMVKFGTDLKEAMD LGTEAAKYVSTLFKHPINLEFEKAYFPYLLINKKRYAGLFWT-N
Schizosaccharomyces SVMVKFGVGTLP EAMKLGEEAANYVSDQFPNPIKLEFEKVIYFPYLLISKKRYAGLFWT-R
*** .:* . : : * * : * * * * * .***** . * ***** .***** : :

Homo PDAHDRMDCKGLEAVRRDNCPLVANLVTASLRRLIDRDPEGAVAHQAQDVISDLLCNRID
Mus SDAHDKMDCKGLEAVRRDNCPLVANLVTSSLRRLIVDRDPDGA VAHAKDVISDLLCNRID
Anolis PDTHDKMDCKGIETVRRDNCPLVANLINTCLQKLLIDRD PAGAVAHAKEVISDLLCNRVD
Xenopus ANTHDKMDCKGIETVRRDNCPLVANLINTCLQKILIDRD PMAVEHAKDVISDLLCNRID
Danio AEHHDKMDCKGIETVRRDNCPLVANLINTCLQNILIDRD PDGAVTHAKEVISDLLCNRID
Drosophila PDTYDKMDCKGIETVRRDNCPLVANLMNSCLQKLLIERDPDGA VAYKQVIADLLCNRID
Caenorhabditis PDVHDKMDCKGLETVRRDNCPLVAKVLGVCLEKLLIERDQQSALDFAKRTISDLLCNKID
Saccharomyces PDKFDKLDQKGLASVRRDSCSLVSI VMNKVLKILIERNV DGALAFVRETINDILHNRVD
Schizosaccharomyces TDTYDKMDSKGIETVRRDNCPLVSYVIDTALRKM LIDQDVEGAQLFTKKVISDLLQNKID
: .:* * ** : ***** . ** : : * .:* * : : : . * .: . * * : * * : *

Homo ISQLVITKELTRAASDYAGKQAHVELAERMRRKRDPGSAPSLGDRVPYV IISAAGVAAAYM
Mus ISQLVITKELTRAAADYAGKQAHVELAERMRRKRDPGSAPSLGDRVPYV IIGAAGVAAAYM
Anolis ISQLVITKELTRTADEYAGKQAHVELAERMRRRDPGSAPNLGDRVPYV IIGAAGVAAAYM
Xenopus ISQLVITKELTRTADEYAGKQAHVELAERMRRKRDPGSAPNLGDRVPYV IIGAAGVAAAYM
Danio ISQLVITKELTRTAQ EYAGKQAHVELAERMRRKRDAGSAPNLGDRVPYV IKAAGVAAAYM
Drosophila ISHLVITKELAK--TDYAAKQAHVELA AKMKKRDPGTAPKLGDRVPYV ICAAAKNTPAYQ
Caenorhabditis ISLLIISKELTKSGDKYQAKQAHVELAARMKKRDAGSAPRLGDRVPYV FVAAAKNVPAYE
Saccharomyces ISKLIISKT LAP---NYTNPQPHAVLAERMKR-REGVGNVGD RVDYV IIGGN--DKLYN
Schizosaccharomyces MSQLVITKALS K--TDYAAKMAHVELAERMRRKRDAGSAPAIGDRVAYV I IKAAGQGDQFYM
:* *:* * * : . * * . * * : * : * * . * : ***** ** : . *

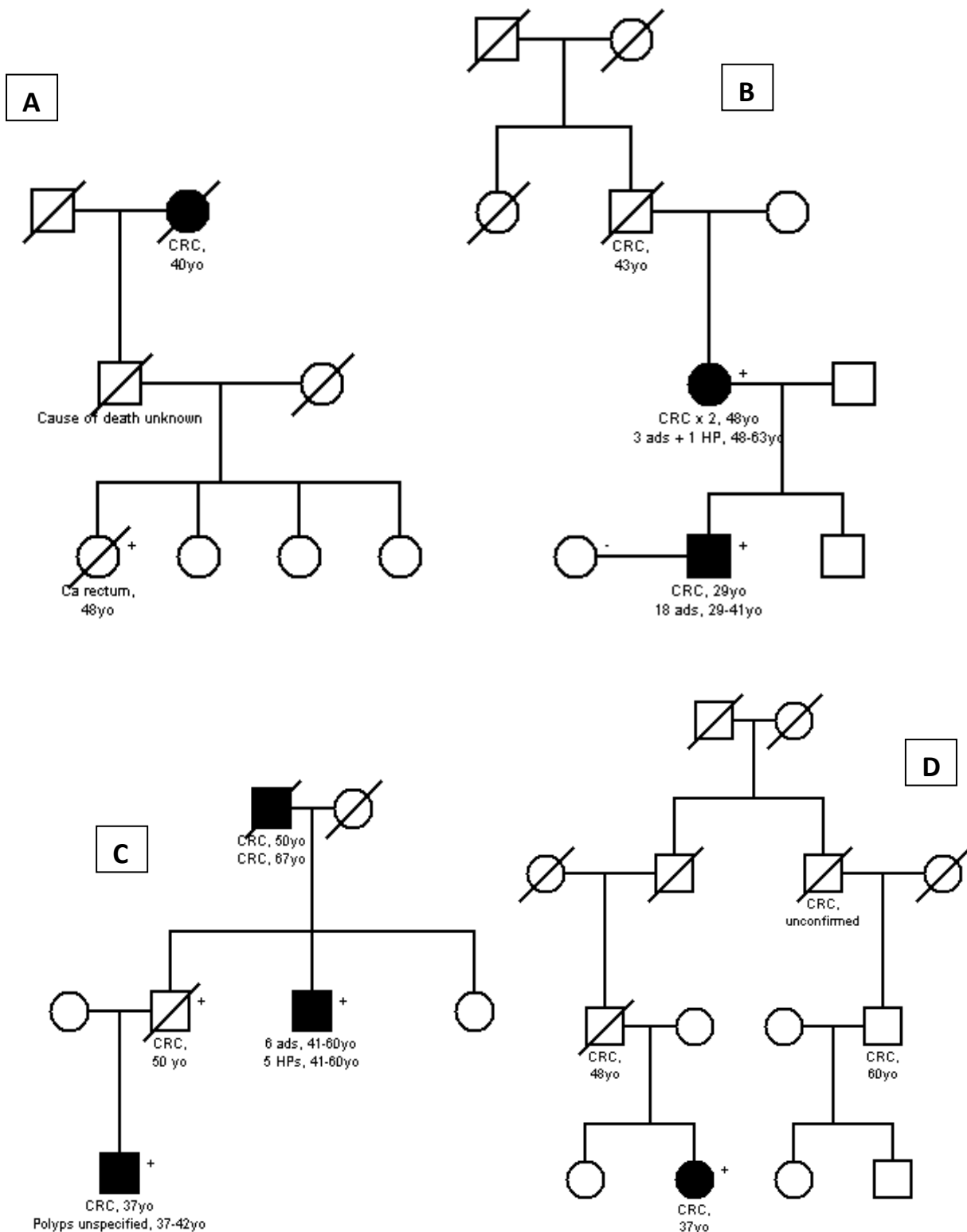
Homo KSEDPLFVLEHSLPIDTQYYLEQQ LAKPLLRIFEPILGEGRAEAVLLRGDHTRCKTVL TG
Mus KSEDPLFVLEHSLPIDTQYYLEQQ LAKPLLRIFEPILGEGRAESVLLRGDHTRCKTVL TS
Anolis KSEDPIYVLENNLPIDTQYYLEQQ LAKPLLRIFEPILGEGKAQNI LLKGEHTRCKTVL TA
Xenopus KSEDPIYVLENNIPIDTQYYLEQQ LAKPLLRIFEPILGESKAESVLLKGEHTRCKTVL TA
Danio KSEDPIYVLENNIPIDTQYYLEQQ LSKPLLRIFEPILGESKAESVLLKGDHTRCKTVL TS
Drosophila KAEDPLYVLENSVPIDATYYLEQQ LSKPLLRIFEPILGDN-AESILLKGEHTRTRTVV TS
Caenorhabditis RAEDPTFVLQNNIPLDTKH YLTNQLAKPLARIFEPILGDR-AEKILVEGEHTRVTVV QS
Saccharomyces RAEDPLFVLENNIQVDSRY YLTNQLQNP IISIVAPIIGDKQ--ANGMFVVKSIKI-NTGS

Schizosaccharomyces	RSEDPIYVLENNIPIDAKYYLENQLSKPLLRIFEPILGEKA--SSLHGDHTRTISMAAP :.*.* :*.:. :* :** :** :* : * .**.* : : :
Homo	KVGGLLAFARRNCCIGCRTVLSH---QGAVCEFCQPRESELYQKEVSHLNALEERFSRL
Mus	KVGGLLAFATKRRNCCIGCRSVIDH---QGAVCKFCQPRESELYQKEVSHLNALEERFSRL
Anolis	KVGGLMAFATKRSTCIGCRAVLNH---HGAVCKFCLSRQSELYQKEVTYLSLSEEEKFSRL
Xenopus	KVGGLMAFAKKRSTCIGCKATLNH---DGAVCNKYCKRESELLQKEISQLSALEEKFSRL
Danio	RVGGLMAFAQKRSTCIGCRAVLKT---DVAVCDFCCKRESELYQKEIAHLSTLEEKFSRL
Drosophila	KVGGLAGFMTKKTSCLGCKSLMPKGYEQACLCPHCEPRMSELYQKEVGAKRELEETFSRL
Caenorhabditis	KVGGLAAFTTKSATCLGCKSVLPRAESENAVCKHCEPKLPTIFASRMNTMHELENHFGRL
Saccharomyces	QKGGMLMSFIKKVEACKSCKGPLRKG--EGPLCSNCLARSSELYIKALYDVRDLEEKYSRL
Schizosaccharomyces	SVGGIMKFAVKVETCLGCKAPIKKG--KTALCENCLNRSSELYQQRQVAQVNDLEVRFARL ** : * : * .* : : . : * * : : : ** : .**

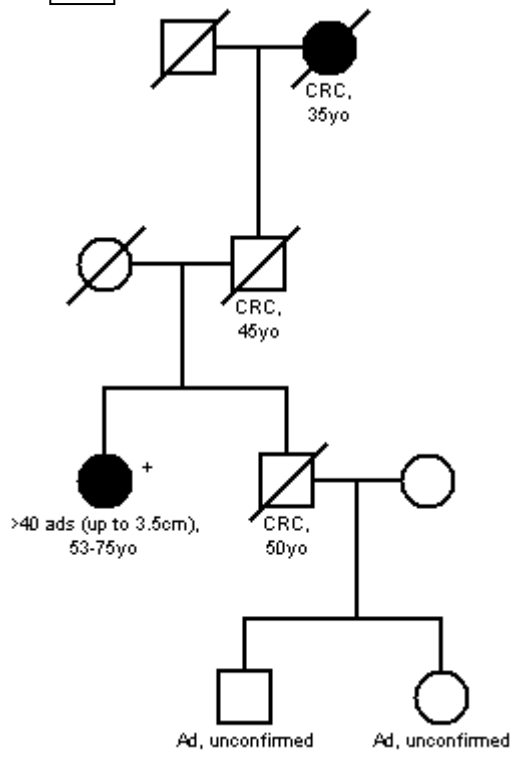
Homo	WTQCQRCQGS LHEDVICTSRDCPIFYMRKKVVRKDLEDQEQLLRRFGPPGPEAW--
Mus	WTQCQRCQGS LHEDVICTSRDCPIFYMRKKVVRKDLEDQERLLQRFPPGPEAW--
Anolis	WTQCQRCQGS LHEDVICTSRDCPIFYMRKKVQKDLDDQEVLSRFGPPT---W--
Xenopus	WTQCQRCQGS LHHEEVLCTSRDCPIFYMRKKVQKDLDDQEKLTFRFGPPA---W--
Danio	WTQCQRCQGS LHEDVICTSRDCPIFYMRKKVQKDLDDQEKLVSRFGW-----
Drosophila	WTECQRCQES LHHEEVICSNRDCPIFYMRQKVRMDLDNQEKRVLRFGLAE---W--
Caenorhabditis	WTECQNCAKTMQDKVNC SARD CPIFYMRKVRNELSEASAVIERFGDPCFQAPTK
Saccharomyces	WTQCQRCAGNLHSEVLC SNKNCDI FYMRVKVKELQEKVEQLSKW-----
Schizosaccharomyces	WTQCQRCQGS MHQDVICTSRDCPIFYMRVIAEHKKLQQSVDLLKRFDEMS---W-- **:*.* .:..* * : :* *:* : .* : :

Supplementary Figure 3. Pedigrees of *POLE* L424V carriers in the Validation Phase.

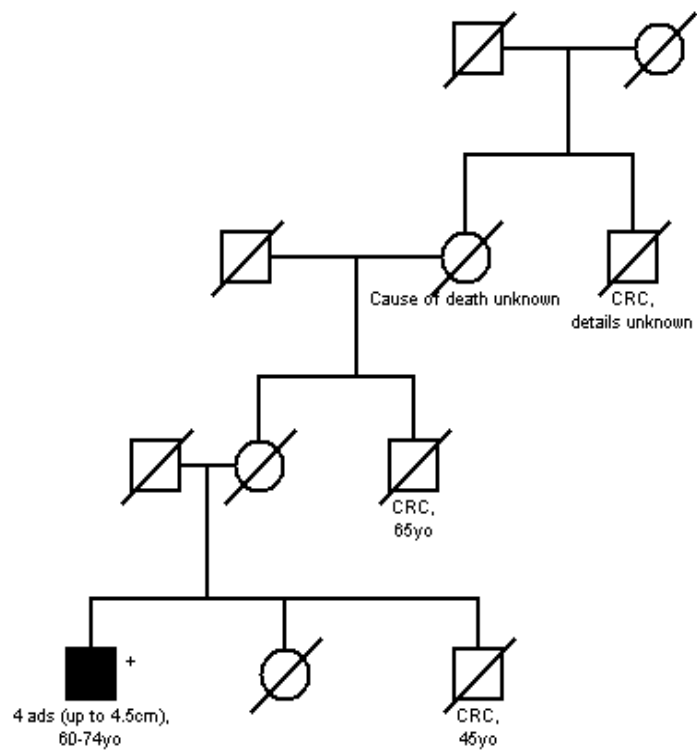
[●]=affected (≥5 colorectal adenomas, ≥1 large (>2cm diameter) adenoma, ≥2 CRCs, or CRC at ≤40 years of age), [+]=heterozygous mutation carrier and [-]=wildtype. For colorectal adenomas (ads), we show the cumulative tumour numbers from age at first presentation or screening colonoscopy to age at last contact. Diameter of the largest adenoma is also given where reported. Hyperplastic polyp (HP) numbers are also shown. For colorectal carcinomas (CRCs), endometrial carcinomas (ECs) and brain tumours, age at first presentation is given. Location of the CRC (Ca colon, caecum, rectum) is also given where reported. One family is not shown, since full family history was not available, although it was known that the proband developed colorectal carcinoma aged 45, and her father was similarly diagnosed at age 28.



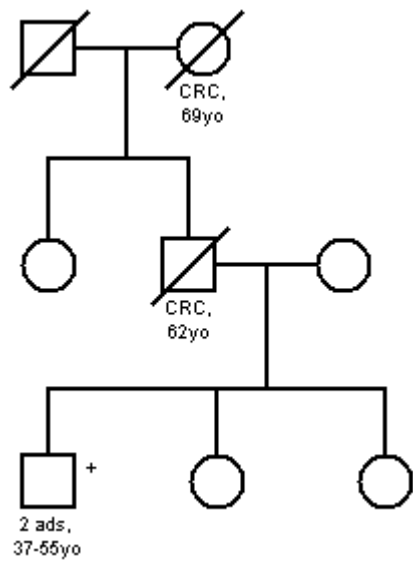
E



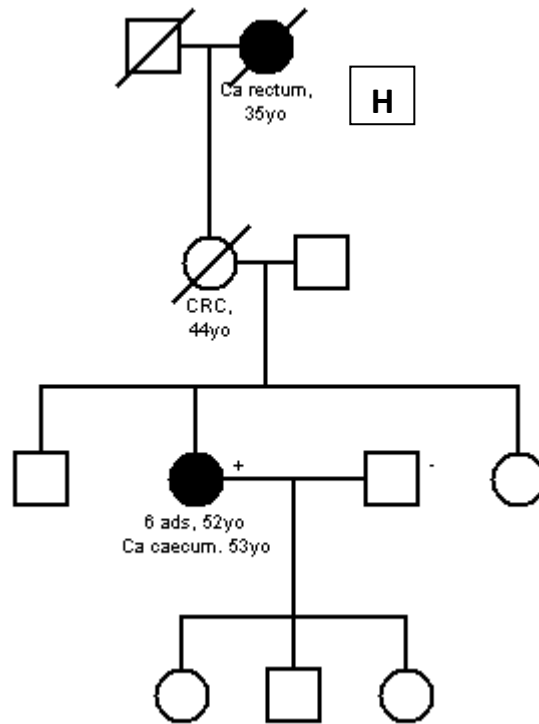
F



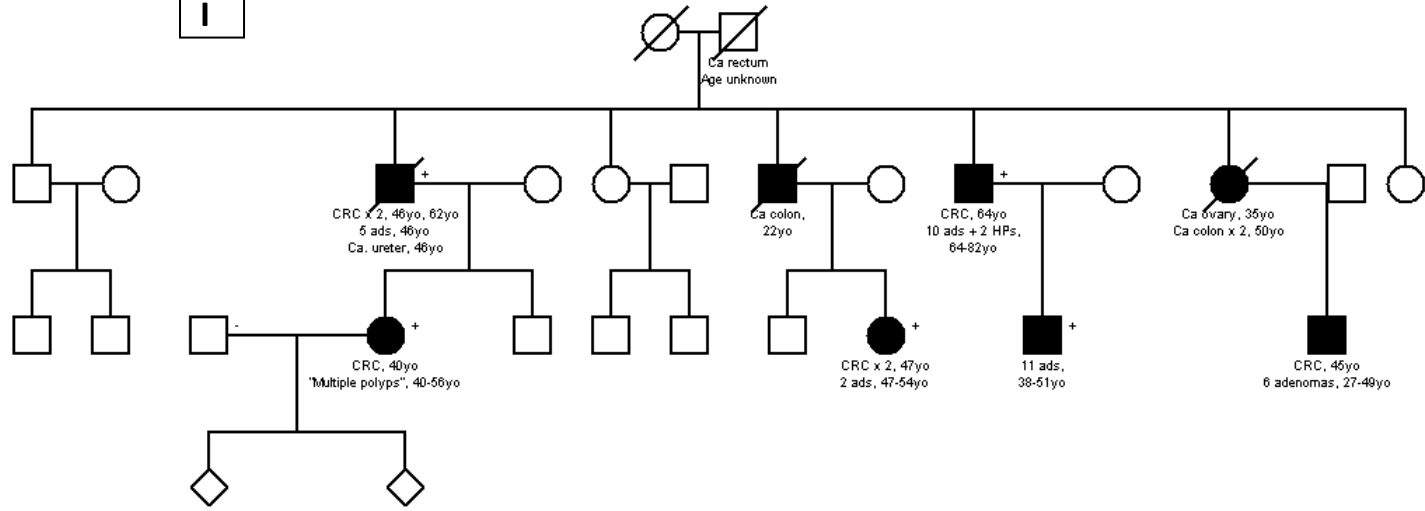
G

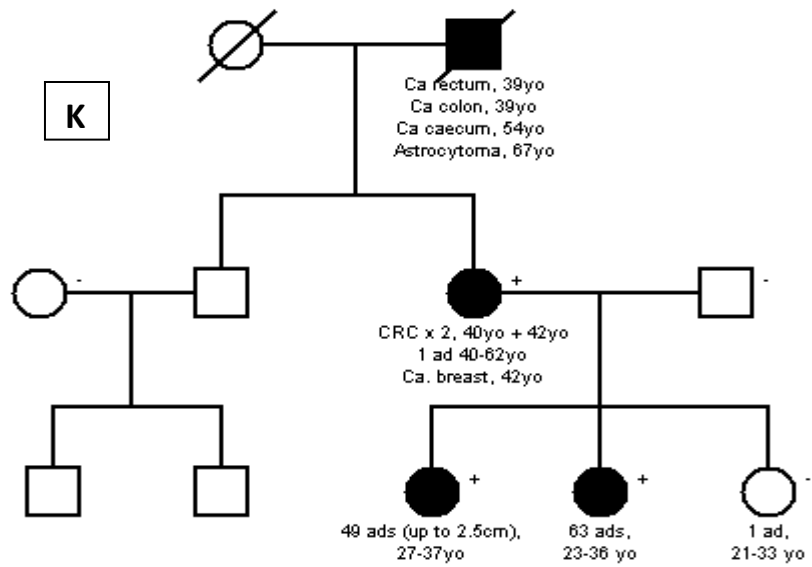
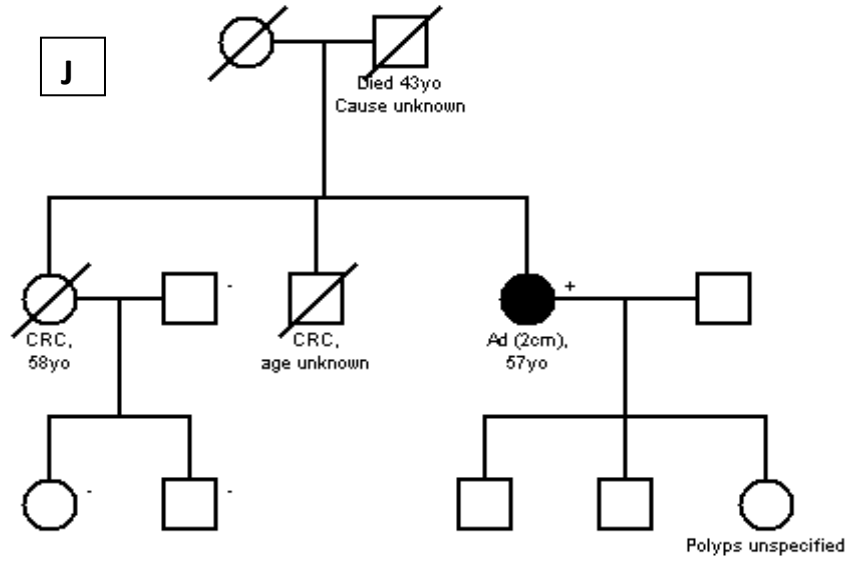


H



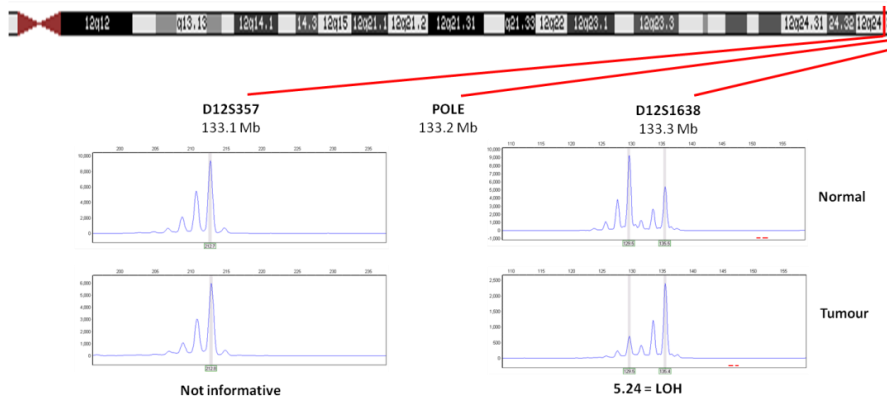
I



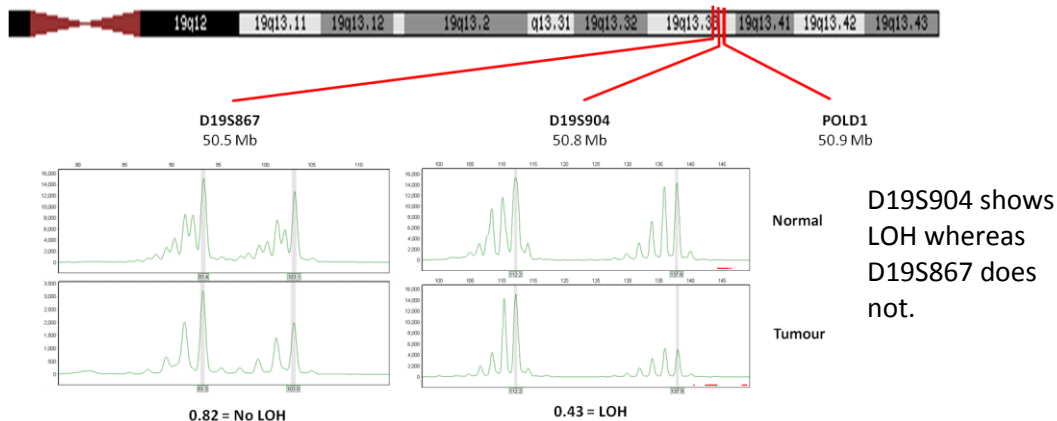


Supplementary Figure 4. Second hits by LOH at *POLE* and *POLD1*.

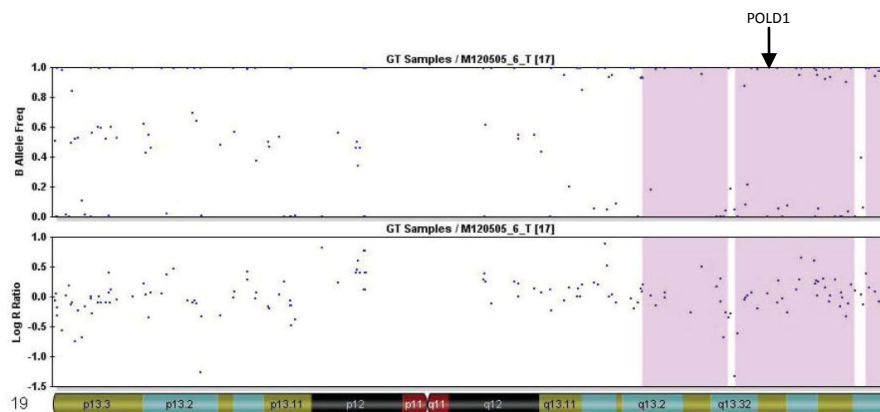
(a) Representative LOH at *POLE* in an adenoma from family SM2702
 Microsatellite markers are shown separated by allelic fragment size. D12S357 is constitutionally homozygous (not informative), but D12S1638 shows LOH.



(b) Representative LOH at *POLD1* in two adenomas from family SM1645, one by microsatellite analysis (upper) and the other by SNP array (lower).



The SNP array data below (tumour only shown) indicate a larger region of LOH extending from the telomere to sub-band q13.2; the position of *POLD1* is arrowed. The LOH is copy-neutral (B allele frequency close to zero or 1 and Log R ratio close to 0). The paired constitutional DNA (not shown) was informative for multiple SNPs between the 19q12 LOH breakpoint and the telomere.



Supplementary Figure 5. POLD1-POLE protein BLAST2 homology search.

Query is POLD1, Subject is POLE. POLE 424 is highlighted in purple and POLD1 478 in red. Exonuclease active site residues from the NCBI Conserved Domain Server are shown in blue.

```

Score = 41.6 bits (96), Expect = 6e-07, Method: Compositional matrix adjust.
Identities = 79/348 (23%), Positives = 129/348 (37%), Gaps = 74/348 (21%)

Query 245 EANVDFEIRFMVDTDIVGCNWLELPAGKYALRLKEKATQCQLEADVLSVSDVVSHPPEGPW 304
          E +V + IR +D I +W Y +R + A ++ D+V P
Sbjct 223 EYDVPYHIRLSIDLKIHVAHW-----YNVRYRGNAFPVEITRR---DDLVERPD---- 268

Query 305 QRIAPLRVLSF[LIEE]AGRKGIFFEPERDPVIQICSLGLRWGEPEPFLRLALTLRPCAPIL 364
          P+ VL+FDIE FP+ E D ++ I + G +T R I+
Sbjct 269 ----PV-VLAF[LIEE]TKLPLKFPDAETDQIMMISYMIDGQG-----YLITNRE---IV 313

Query 365 GAKVQSYE-----KEEDLLQAWSTFIRIMDPDVTG[YN]IQ[NEE]LPYLI 407
          ++ +E E L+Q W ++ P ++ YN FD P++
Sbjct 314 SEDIEDFEFTPKPEYEGPFCVFNEPDEAHLIQRWFEHVQETKPTIMVT[YN]G[EED]WPFVE 373

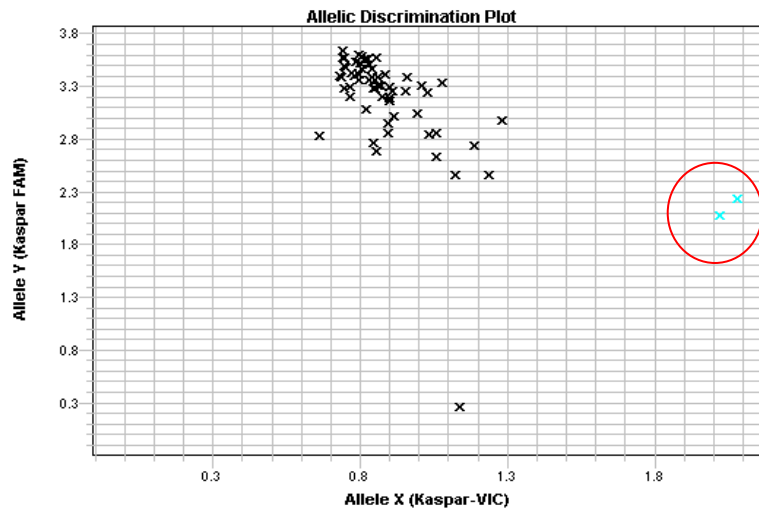
Query 408 SRAQTLKVQTFPFLGRVAGLCSNIRDSSFQSKQTGRRDTKVVSMVGRVQMDMLQVLLREY 467
          +RA V GL S ++ FQ G + MD L+ + R+
Sbjct 374 ARAA-----VHGL-SMQQEIGFQKDSQGEYKAPQC-----IHMDCLRWVKRDS 415

Query 468 KL--RSY[TL]NAV[S]FHLGEQKEDVQHSIITDLQNGNDQTRRRLAV[Y]CLK[D]AYLPLRLLER 525
          L S+ L A + LG ++ + + QT LA Y + DA L
Sbjct 416 YLPVGSH[N]KAAAKAKLGYDPVELDPEDMCRMATEQPQT---LAT[YS]VS[D]AVATYYL--- 469

Query 526 LMVLVNAVEMARVTGVPL--SYLLSRGQQVKVVSQLLRQAMHEGLLMP 571
          M V+ A T +P+ +L +G + L+ QA H ++ P
Sbjct 470 YMKYVHPFIFALCTIIPMEPDEVLRKGSGLCEALLMVQAFHANIIFP 517
    
```

Supplementary Figure 6. KASPar genotype clusters for POLE L424V

The blue crosses are mutants, subsequently confirmed by direct sequencing. The sample at the bottom is a blank. The remaining samples are wildtype.



Supplementary Table 1. Features of the 15 probands who underwent whole-genome sequencing.

In the family history field, S=son, D=daughter, M=mother, F=father, Si=sister, B=brother, N= niece or nephew, U=uncle, A=aunt, C=first cousin, GM=grandmother, GD=granddaughter, ads=adenomas, HPs=hyperplastic polyps, yo=years old at initial presentation. Individuals with 2 or fewer adenomas or hyperplastic polyps are not shown. Individuals shown in bold were additional relatives who underwent whole-genome sequencing. For families SM2702, SM1645 and SM1412, the clinical details, including the features of the additional relatives sequenced in those kindreds, are shown in Figure 1

Family number	Sex	Age at presentation	Age at last investigation	Adenomas to date	CRC	Mutations found in known CRC genes	Family history of colorectal tumours
SM2702	F	23	30	68	No		See Figure 1(a)
SM1645	F	53	66	17	No		See Figure 1(b)
CAM13042	F	57	69	12	Yes		Si 3 ads @ 61yo; B 2 CRCs @ 60 yo; N 7 ads @ 35yo
OX39787	F	33	33	60	Yes		No
SM1613	M	46	48	11	No		C 4 ads @ 46yo ; C 4 ads 65yo' U CRC 70yo; U CRC 70yo; A CRC 82yo
OX42383	M	50	53	9	Yes		S 6 ads @ 42yo; S CRC @ 44yo; S CRC @ 54 yo
SM3708	M	60	60	16	No		B 10 ads @ 66yo; N 5 ads + 7 HPs, 37yo; B CRC @ 72yo
BHAM12930	F	47	47	40	Yes		D 4 ads @ 32yo; S 3 ads @ 29yo
BHAM21930	M	40	40	10	No		F CRC+15 ads @ 70yo ; B 3 ads @ 53yo; Si 2 HPs @ 48yo; Si 1 ad @ 39yo; B 3 HPs @ 39yo
SM1412	F	41	53	19	Yes		See Figure 1(c)
DUB2183	F	39	39	10	No	MSH6 R911X	S CRC @ 38yo; U 12 ads + 3 HPs @ 59yo
SM2349	M	29	59	10	No		B CRC + 4 ads + 16 HPs @ 63yo
ManP0100251	F	60	60	17	Yes	APC c.423 -2 A>G	D CRC + 7 ads @ 53yo; GD 3 ads @ 31 yo
OX41405	M	24	24	36	Yes		No
SM3162	F	32	32	>20	No		C 80 HPs @ 29 yo; GM CRC @ 76yo

Supplementary Table 2. Coding variants within regions shared in linkage analysis of SM2702.

Data are LOD scores derived from non-parametric (affecteds-only) analysis of Affymetrix 10K array data using MERLIN (see Methods). cM=centiMorgans, maxLOD=maximum log(odds), B37=human genome build 37.

Chr	cM LOD>1.2	MaxLOD	B37 flanking base 1	B37 flanking base 2	Gene	Variant	Gene	Variant
1	125-165	1.78	112,380,534	166,550,501				
2	195-220	1.78	193,440,950	217,137,980				
3	125-140	1.77	121,838,319	128,726,030	STXBP5L	A791T		
5	165-190	1.78	167,563,585	174,609,731				
9	30	1.71	5,267,043	14,181,653				
10	120-145	1.78	98,981,914	120,110,549	FBXW4	A19G	TECTB	I254T
12	152-170	1.59	127,823,984	133,851,895	POLE	L424V		
19	20-40	1.57	6,828,040	11,797,308	C19orf71	R38C	ELAVL3	D296N

Supplementary Table 3. Details of somatic changes in POLE L424V mutation carriers' colorectal tumours

The rows show individual tumours analysed for LOH, MSI, *KRAS* and *BRAF* in almost all cases, and for other CRC driver mutations in a subset of tumours. Patient details are shown and can be matched to the pedigrees in Figure 1 (numbers) or Supplementary Figure 3 (letters) using the clinical features. Molecular results are shown for: LOH at *POLE*; *APC* mutations (protein-truncating between codons 167 and 1580); *CTNNB1* mutations (codons 32, 33, 34, 37, 41, 45); *KRAS* mutations (codons 12, 13, 61, 146); *BRAF* mutations (codon 600); *PIK3CA* mutations (codons 345, 420, 542, 545, 546, 1025, 1047); and *FBXW7/CDC4* mutations (codons 465, 479, 505, 582 and protein-truncating). Mutation details are shown as protein change and location, and base change. All specific mutations were heterozygous except where LOH is shown. WT=wild type; dark shaded cells=not tested or failed. Aneuploidy was taken as indicative of chromosomal instability (CIN).

Family no.	Patient no.	Phenotype	Block code	Tumour type	LOH POLE	MSI	PLOIDY	APC	CTNNB1	KRAS	BRAF	PIK3CA	FBXW7	
2702	1	CRC + 5 ads, 43yo	A	Adenoma	No LOH	MSS		Q1338X (C>T)	WT	WT	WT	WT	R465C (C>T)	
			B	Adenoma	No LOH	MSS				WT	WT			
2702	2	9 ads + 1 HP, 38-51yo	A	Adenoma	No LOH	MSS			WT	WT				
2702	3	CRC, 40yo; 27 ads, 40-54yo	B	Adenoma	No LOH	MSS				WT	WT			
			C	Adenoma	No LOH	MSS				WT	WT			
			D	Adenoma	No LOH	MSS					WT			
			E	Adenoma	No LOH	MSS								WT
			F	Adenoma	LOH	MSS				WT	WT			
			G	Adenoma	LOH	MSS				WT	WT			
			H	Adenoma	No LOH	MSS				WT	WT			
			I	Adenoma	LOH	MSS				WT	WT			
			J	Adenoma	No LOH	MSS				WT	WT			
			K	Adenoma	LOH	MSS					WT			
			L	Adenoma	No LOH	MSS				WT	WT			
			M	Adenoma	No LOH	MSS				WT	WT			
			N	Adenoma	No LOH	MSS				WT	WT			
			O	Adenoma	No LOH	MSS				G12D (G>A)	WT			
			P	Adenoma	No LOH	MSS				WT	WT			
			Q	Adenoma	No LOH	MSS				WT	WT			
			R	Adenoma	No LOH	MSS				WT	WT			
			A	Carcinoma	No LOH	MSS	Aneuploid	R1114X (C>T), Q1338X (C>T)	WT	A146T (A>C)	WT	WT	WT	
A	1	Ca rectum, 48yo	A	Carcinoma	LOH	MSS	Aneuploid			WT	WT			
I	1	CRC, 40yo; Multiple polyps, 40-56yo	A	Adenoma	No LOH	MSS		Q1338X (C>T)	WT	A146T (A>C)	WT	WT	R479Q (G>A)	
I	2	CRC, 64yo; 10 ads + 2 HPs, 64-82yo	A	Adenoma	LOH	MSS		Q1338X (C>T)	WT	WT	WT	WT	WT	
			B	Adenoma	No LOH	MSS				WT	WT			
			C	Adenoma	No LOH	MSS				G12D (G>A)	WT			
			D	Adenoma	LOH	MSS				G12D (G>A)	WT			
			E	Adenoma	LOH	MSS					WT			
			F	Adenoma	No LOH	MSS					WT			
			G	Adenoma	No LOH	MSS				G12D (G>A)	WT			
			H	Adenoma	No LOH	MSS				WT	WT			
			I	Adenoma	No LOH	MSS					WT			
			J	Adenoma	LOH	MSS				WT	WT			
F	1	4 ads (up to 4.5cm), 60-74yo	A	Adenoma	No LOH	MSS				WT	WT			
			B	Adenoma	LOH	MSS				WT	WT			
			C	Adenoma	No LOH	MSS				WT	WT			
K	1	49 ads (up to 2.5cm), 27-37yo	A	Adenoma	No LOH	MSS		R1114X (C>T)	WT	WT	WT	WT	WT	
C	1	CRC, 50yo	A	Carcinoma	No LOH	MSS				WT	WT			
H	1	6 ads, 52yo; Ca caecum, 53yo	B	Adenoma	No LOH	MSS				WT	WT			

Supplementary Table 4. Coding variants within regions shared in linkage analysis of SM1645.
 Legend is as for Supplementary Table 2. On testing two additional affected family members, only *POLD1* S478N was shared by all affecteds in the family.

Chr	cM LOD>1.2	MaxLOD	B37 flanking base 1	B37 flanking base 2	Gene	Variant
1	155-180	1.17	161,849,180	179,520,151		
1	250-end	1.32	239,785,165	248,793,746		
2	55-65	1.44	33,916,066	37,949,251		
2	85-100	1.48	69,747,817	73,374,865		
2	130-150	1.15	121,367,230	133,742,214		
3	160-185	1.33	149,385,539	173,189,273	ECT2	S500N
4	190-end	1.28	187,390,861	190,716,184		
6	145-175	1.42	145,430,900	161,369,806	IGF2R	R292W
7	60-75	1.54	37,563,478	54,248,108		
18	95-115	1.37	65,958,625	72,538,424	RTTN	P2135L
19	75-end	1.36	46,650,324	58,745,931	POLD1	S478N

Supplementary Table 5. Details of somatic changes in *POLD1* S478N mutation carriers' colorectal tumours

Legend is as for Supplementary Table 3, except that LOH is at *POLD1*.

Family no.	Patient no.	Phenotype	Block code	Tumour type	LOH <i>POLD1</i>	MSI	PLOIDY	APC	CTNNB1	KRAS	BRAF	PIK3CA	FBXW7
1412	1	CRC, 32yo; 45 ads (up to 3.5cm) + HPs, 32-52yo	A	Carcinoma	No LOH	MSS	Aneuploid	WT	WT	WT	WT	WT	WT
1645	1	Ca splenic flexure, 28yo; 28 ads, 28-35yo	A	Carcinoma	LOH	MSS		E1309X (C>T)	WT	Q61H (A>C)	WT	WT	WT
			C	Adenoma	LOH								
			B	Adenoma	LOH	MSS			WT	WT	WT		
1645	2	10 ads (up to 4cm), 43-65yo	A	Adenoma	LOH	MSS		R1114X (C>T), Q1367X (C>T)	WT		V600E (T>A)	WT	WT
			B	Adenoma	LOH	MSS		Q1338X (C>T)	WT	G13D (G>A)	WT	WT	WT
			C	Adenoma	No LOH	MSS		Q1338X (C>T)	WT	G13D (G>A)	WT	WT	R465H (G>A)
			D	Adenoma	No LOH	MSS		Q1338X (C>T)	WT	G13D (G>A)	V600E (T>A)	WT	R465C (C>T)
			E	Adenoma	LOH	MSS			WT	G13D (G>A)	WT		
1645	3	6 ads and "multiple" HPs, 29-39yo	A	Adenoma	No LOH								
1645	4	EC, 52yo; 28 ads + 2HPs, 52-68yo	A	Adenoma	No LOH								
			B	Adenoma	No LOH								
			C	Adenoma	No LOH								
			D	Adenoma	No LOH								
			E	Adenoma	No LOH								
			F	Adenoma	No LOH								
			G	Adenoma	No LOH								
			H	Adenoma	No LOH								
1645	5	EC, 45yo; 17 ads, 53-65yo	A	Adenoma	No LOH								
			B	Adenoma	No LOH								
			C	Adenoma	No LOH								
			D	Adenoma	No LOH								
			E	Adenoma	No LOH								

Supplementary Table 6. Mutator phenotype in *S. pombe* carrying the equivalent of the human POLD1 S478N mutation.

(a) Fission yeast strains used are shown.

Strain no.	Genotype	source
2840	<i>h⁻ ade6-485</i>	O. Fleck
3176	<i>h⁻ pol3⁺::kanMX6 ade6-485</i>	
3177	<i>h⁻ pol3-C462N::kanMX6 ade6-485</i>	
3178	<i>h⁻ pol3-C462S::kanMX6 ade6-485</i>	

(b) Clustal sequence alignment of the catalytic subunits of human, *S. pombe* and *S. cerevisiae* Pol δ and Pol ϵ in the region between the ExoII and III motifs

HumanPOLE1	391KDSQGEYKAP	QCIHMDCLRW	VKRDSYLPVG	SHN L KAAAKA	KLGYDPVELD
SpPol2	392RDAEDEYKSS	YCSHMDAFRW	VKRDSYLPQG	SQGLKAVTVS	KLGYNPIELD
ScPol2	406PDAEGEYKSS	YCSHMDCFRW	VKRDSYLPQG	SQGLKAVTQS	KLGYNPIELD
humanPOLD1	443RRDTKVSMV	GRVQMDMLQV	LLREYKL--R	SYTLNAV S FH	FLGEQKEDVQ
SpPol3	427TRESKTTTIP	GRLQLDMLQV	MQRDFKL--R	SYSLNAVCSQ	FLGEQKEDVH
ScPol3	448TRETKNVNID	GRLQL D LLQF	IQREYKL--R	SYT L NAVSAH	FL G EQKEDVH
		*	* * *	* * *	**

POLE1 L424V and POLD1 S478N are shown (bold, yellow), together with residues which cause a mutator phenotype in *S. cerevisiae* Pol3 (L479S, D463G/G488D (Murphy et al. Genome, 49, 403-410, 2006); red). Asterisks show residues completely conserved in this comparison.

Note that in the RB69 DNA polymerase structure, L290 (equivalent to *S. cerevisiae* L479 and *H. sapiens* L424) is packed together with W216, which is in the Exo II domain; this forms a platform for ss primer DNA (Murphy et al. Genome, 49, 403-410, 2006; Hogg et al. EMBO J, 23, 1483-1493, 2004). These residues appear to move between the editing and polymerizing conformations, so they may be involved in moving the primer from the polymerase to the exonuclease active site.

(c) PCR primers used to amplify *pol3* segments were (5'-3'):

1075 ttggatccttgaatgcgaggtcgcaaaggcg;

1076 tggcgcgccccatcttttcaccaggacatttcatcaaatc;

1077 agagttaactgcatttaaagagtaagatcgaagcttaaaatcac;

1078 gcagttaactctcagtttctaggcgagcaaaaagaag;

1079 agaagagactgcatttaaagagtaagatcgaagcttaaaatcac; and

1080 gcagtctcttctcagtttctaggcgagcaaaaagaag.

Supplementary Table 7. Features of POLE-mutant and POLD1-mutant TCGA colorectal cancers.

The specific somatic mutations shown are the same as those in Supplementary Tables 3 and 5. All mutations are heterozygous. MSI is shown divided into stable (MSS), low (MSI-L) and high (MSI-H) as per the original TCGA classification, but since there is no clear evidence that MSI-L and MSS are distinct groups, we have combined them into an MS-stable group for our analysis. Note that all *POLD1* mutations are in MS-unstable cancers, although one occurs in the exonuclease domain and might be functionally important. Cat=catalytic domain, Exo=exonuclease domain, DFU=putative C-terminal DNA-binding domain, None=no specific domain, Loss=protein-truncating mutation presumed to be null. Mutations affecting putative active site residues are shown in bold. By bioinformatic and structural analysis (details not shown), we found no good evidence for the functional importance of the missense non-exonuclease domain mutations (A189T, K777N, L1255V) that are present in cancers which also have an exonuclease domain mutation.

No.	Tumour	Hyper-mutated?	MSI	MLH1 methylated?	Gene	AA Change	POL domain	Structural assessment (exonuclease domain mutants)	APC mutations	CTNNB1 mutation	KRAS mutation	BRAF mutation	PIK3CA mutation	FBXW7 mutation
1	TCGA-AA-3516	Yes	MSI-H	Yes	POLD1	p.A864T	Cat		None found	None found	None found	V600E (T>A)	None found	None found
2	TCGA-AA-3710	Yes	MSI-H	Yes	POLD1	p.A145D	None		None found	None found	None found	None found	None found	None found
3	TCGA-AA-3947	Yes	MSI-H	Yes	POLD1	p.P787L	Cat		E847F5	None found	None found	V600E (T>A)	H1047R (A>G)	None found
4	TCGA-AA-3949	Yes	MSI-H	Yes	POLD1	p.Q461H	Exo	DNA-binding interface	None found	None found	None found	V600E (T>A)	None found	R367X (C>T)
5	TCGA-AA-A00J	Yes	MSI-H	Yes	POLD1	p.R808H	Cat		R554X (C>T)	None found	None found	V600E (T>A)	None found	None found
5	TCGA-AA-A00J	Yes	MSI-H	Yes	POLE	p.A2056T	None							
6	TCGA-AA-3518	Yes	MSI-H	Yes	POLE	p.V1368M	None		None found	None found	None found	None found	None found	R465C (C>T)
7	TCGA-AA-3525	Yes	MSI-H	Yes	POLE	p.A2213V	None		None found	None found	None found	V600E (T>A)	None found	None found
7	TCGA-AA-3525	Yes	MSI-H	Yes	POLE	p.K1008N	Cat							
7	TCGA-AA-3525	Yes	MSI-H	Yes	POLE	p.R762W	Cat							
8	TCGA-AA-3555	Yes	MSS	No	POLE	p.P286H	Exo	DNA-binding interface	R499X (C>T), R1450X (C>T)	None found	A146T (G>A)	None found	None found	None found
9	TCGA-AA-3678	No	MSS	No	POLE	p.D1752N	DFU		E1309F5	None found	None found	None found	None found	None found
10	TCGA-AA-3710	Yes	MSI-H	Yes	POLE	p.P1421S	None		None found	None found	None found	None found	None found	None found
11	TCGA-AA-3864	Yes	MSI-H	No	POLE	p.R231H	None		R348X (C>T), R564X (C>T), R1432X (C>T)	None found	None found	None found	None found	None found
12	TCGA-AA-3977	Yes	MSS	No	POLE	p.K777N	Cat		R1114X (C>T), E1309X (G>T)	None found	None found	None found	None found	None found
12	TCGA-AA-3977	Yes	MSS	No	POLE	p.F367S	Exo	DNA-binding interface						
13	TCGA-AA-3984	Yes	MSS	No	POLE	p.V411L	Exo	DNA-binding interface	R1114X (C>T)	None found	None found	None found	None found	None found
14	TCGA-AA-A00N	Yes	MSI-L	No	POLE	p.L1255V	None		S1281X (C>A), E1408X (G>T)	None found	G13D (G>A)	None found	None found	None found
14	TCGA-AA-A00N	Yes	MSI-L	No	POLE	p.V411L	Exo	DNA-binding interface						
15	TCGA-AA-A010	Yes	MSI-L	No	POLE	p.P436R	Exo	In disordered loop, ordered on DNA binding	R1450X (C>T)	None found	None found	None found	None found	None found
15	TCGA-AA-A010	Yes	MSI-L	No	POLE	p.A189T	None							
16	TCGA-AG-3892	Yes	MSS	No	POLE	p.S459F	Exo	DNA-binding interface	R1114X (C>T)	None found	None found	None found	None found	None found
17	TCGA-AG-A002	Yes	MSS	No	POLE	p.S459F	Exo	DNA-binding interface	R1450X (C>T), E1538X (G>T)	None found	None found	None found	None found	None found
17	TCGA-AG-A002	Yes	MSS	No	POLE	p.R150X	Loss							
18	TCGA-AG-A01W	No	MSS		POLE	p.D2013N	None		R876X (C>T)	None found	G13D (G>A)	None found	None found	None found

Supplementary Table 8. (a) Genome-wide substitution mutation spectrum in TCGA colorectal cancers and (b) Ion Torrent custom 150-gene panel mutation spectrum in tumours from *POLE* L424V and *POLD1* S478N carriers

(a) The table shows the proportion and mean number of the 6 types of somatic base substitution mutation in cancers according to their *POLE* mutation type. Cancers with one mutation in Exo domain and no other detected mutation were placed in the “exonuclease domain-mutant” (EDM) category. Cancers with one mutation in the exonuclease domain and another one elsewhere were also placed in the EDM category. Note that the EDM cancers have a higher mutation frequency than either of the other cancer types, and the mutation spectrum is shifted towards G:C>T:A changes. All but two cancers with “other” *POLE* mutation(s) were MS-unstable owing to defective mismatch repair. Whilst these cancers had a mutation frequency higher than that of the *POLE*-wildtype cancers, this was accounted for solely by the MSI, because the substitution frequency and spectrum were very similar between MS-unstable *POLE*-mutant and *POLE*-wildtype cancers (details not shown). Overall, the data show that the “other” *POLE* mutations appear to have no discernible effect of mutation frequencies, in contrast to the exonuclease domain mutations.

Mutation type	POLE exonuclease domain-mutant		Other POLE-mutant		POLE-wildtype	
	Proportion of all substitution mutations	Mean mutations per cancer	Proportion of all substitution mutations	Mean mutations per cancer	Proportion of all substitution mutations	Mean mutations per cancer
C>T, G>A	51.1%	1881	69.8%	577	60.3%	54
G>T, C>A	35.2%	1297	11.4%	94	14.2%	13
A>T, T>A	0.5%	19	2.2%	18	5.1%	5
C>G, G>C	0.2%	7	1.0%	8	6.7%	6
A>G, T>C	5.6%	207	13.2%	109	9.0%	8
A>C, T>G	7.3%	268	2.4%	20	4.7%	4
Total		3678		827		90

(b) The table is as above, but shows data from individual tumours from *POLE* L424V and *POLD1* S478N carriers. The tumour IDs correspond to those in Supplementary Tables 3 and 5. By comparison with the whole-exome TCGA data set obtained on the Illumina and SOLID platforms, the 150-gene Ion Torrent data available from these tumours was necessarily much more limited. We emphasise that comparisons across different sequencing platforms must be made with caution, but we note that our patients' tumours did show similar mutation spectra to that in the *POLE* EDM cancers.

	POLD1 (SM1645, patient 1, Block A)		POLD1 (SM1645, Patient 1, Block B)		POLD1 (SM1412, Patient 1, Block A)		POLE (SM2702, Patient 2, Block A)	
Mutation type	Proportion of all substitution mutations	Mean mutations per cancer	Proportion of all substitution mutations	Mean mutations per cancer	Proportion of all substitution mutations	Mean mutations per cancer	Proportion of all substitution mutations	Mean mutations per cancer
C>T, G>A	27.9%	24	34.4%	11	27.4%	59	38.3%	18
G>T, C>A	20.9%	18	31.3%	10	44.2%	95	23.4%	11
A>T, T>A	10.5%	9	3.1%	1	4.7%	10	4.3%	2
C>G, G>C	9.3%	8	9.4%	3	3.3%	7	8.5%	4
A>G, T>C	19.8%	17	9.4%	3	8.4%	18	10.6%	5
A>C, T>G	11.6%	10	12.5%	4	12.1%	26	14.9%	7
Total		86		32		215		47

Supplementary Note: Additional Methods

Linkage analysis

Some affected members of SM1645 and SM2702 had previously been genotyped using Affymetrix Human Mapping 10K Xba arrays (Papaemmanuil, E. et al. Eur J Hum Genet 16, 1477-86; 2008). Genotypes were called using the dChip SNP software. Haplotype sharing and linkage analysis were performed using the MERLIN program (Abecasis, G.R. et al. Nat Genet 30, 97-101; 2002).

Whole-genome sequencing: library preparation and sequencing

Samples were quantified using the Qubit system (Invitrogen) and sample integrity was determined by 1% agarose gel electrophoresis. 200 ng of DNA were fragmented using the Covaris S2 system and libraries constructed using the NEBNext DNA Sample Prep Master Mix Set 1 Kit (NEB). Ligation of adaptors was performed using 6 µl of the Illumina Multiplexing Sample Preparation Oligonucleotide Kit. Libraries were size-selected using 2% gel electrophoresis and the distribution of fragments in the purified fraction was determined using TapeStation 1DK system (Agilent/Lab901). Each library was PCR-enriched using the following custom primers:

Multiplex PCR primer 1.0: 5'-

AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

Index primer: 5'-

CAAGCAGAAGACGGCATACGAGAT[INDEX]CAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

Indexes were 8bp long and part of an indexing system developed in-house. Four independent PCR reactions per sample were prepared using 25% volume of the pre-PCR library each. After 8 cycles of PCR (cycling conditions as per Illumina recommendations) the four reactions were pooled and purified with AmpureXp beads. The final size distribution was determined using a TapeStation 1DK system (Agilent/Lab901). The concentration of each library was determined by the Agilent qPCR Library Quantification kit. Whole Genome Sequencing was performed on the Illumina HiSeq2000 as 100bp paired end reads, using v3 clustering and sequencing chemistry. A PhiX control was spiked into the library. We ran 2 lanes of the original library at 21.5pM and 23.0pM, and then topped up coverage as required by multiplexing over several lanes at 18.0pM and 18.5pM.

Whole Genome Sequencing: Data Analysis

WGS reads were mapped to the human reference genome (GRCh37/hg19) using STAMPHY and duplicate reads removed using Picard (<http://www.picard.sourceforge.net/>). The mean coverage of mapped reads across the genome was 44.3x with 98.8% of bases covered at ≥15x. Identification of variant sites and alleles was performed with Platypus. Variants were only called if they were assigned a sufficiently high posterior probability (phred score of 5). Additional filters were used to remove variants in low quality reads and variant present only on one strand. Illumina Omni1 SNP array data were used for comparison and 99.4% of the SNP calls were also reported by the WGS. Variants were then annotated using the ANNOVAR software package. The following ANNOVAR databases (with respect to human genome hg19) were used: RefSeq gene models; dbSNP (Build 132); 1000 genomes allelic frequencies (November 2011); UCSC segmental duplication scores; and UCSC 46 species conservation scores. Candidate variants were annotated with predictions of functional importance from SIFT, PolyPhen2, PhyloP and MutT. We also screened for known associations with diseases using OMIM, HGMD Professional and GeneCards.

Microsatellite instability and ploidy analysis

Microsatellite instability was determined using a panel of 3 markers (BAT25, BAT26, D2S123). Tumours were classified as MSI+ if they had one or more unstable markers and MSI- if all markers were stable. Ploidy analysis was performed using the automated image-based cytometry system Ploidy Work Station (Room4, East Sussex, UK). Briefly, diploids had one G0/G1 peak (2c), the number of nuclei in the G2 peak (4c) did not exceed 10% of the total number of nuclei and the number of nuclei with a DNA content exceeding 5c did not exceed 1%; tetraploids had a peak in the 4c position together with a G2 peak in the 8c position or the fraction of nuclei in the tetraploid region (4c) exceeded 10% of the total number of nuclei; polyploids had a peak in the 8c position together with a G2 peak in the 16c position; aneuploids had noneuploid peaks or the number of nuclei with a DNA content exceeding 5c/9c, not representing euploid populations, exceeded 1%. Blocks of only normal tissue were blindly analysed and all of them were scored as diploid. Diploids were classified as chromosomally-stable (CIN-) and tetraploids, polyploids and aneuploids as CIN+.