## Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome

Deborah L. Nagle ${ }^{1}$, Mohammad A. Karim ${ }^{2}$, Elizabeth A. Woolf!, Lisa Holmgren ${ }^{1}$, Peer Bork3, Donald J. Misumi1, Sonja H. McGrail ${ }^{1}$, Barry J. Dussault, Jr. ${ }^{1}$, Charles M. Perou4, Raymond E. BoissyS, Geoffrey M. Duyk ${ }^{1}$, Richard A. Spritz ${ }^{2}$ \& Karen J. Moore ${ }^{1}$

Chediak-Higashi syndrome (CHS\} is a rare, autosomal recessive disorder characterized by hypopigmentation, severe immunologic deficiency with neutropenia and lack of natural killer (NK\} cells, a bleeding tendency and neurologic abnormalities ${ }^{1}-4$. Most patients die in childhood. The CHS hallmark is the occurrence of giant inclusion bodies and organelles in a variety of cell types, and protein sorting defects intothese organelles5-8. Similar abnormalities occur in the beige mouse ${ }^{6}, 7.9$, the proposed model for human CHS. Two groups have recently reported the identification of the beige gene ${ }^{1}, 415$, however the two cDNAs were not at all similar. Here we describe the sequence of a human cDNA homologous to mouse beige, identify pathologic mutations and clarify the discrepancies of the previous reports. Analysis of the CHS polypeptide demonstrates that its modular architecture is similar to the yeast vacuolar sorting protein, VPS15.

Multiple rounds of screening human cDNA libraries with mouse beige probes and subsequently with probes
derived from those newly isolated cDNA clones were performed to obtain 27 human cDNA clones. Sequence assembly of these yielded a sequence of $13,449 \mathrm{bp}$ for the human beige cDNA homologue, CHS, that mapped to chromosome lq43 by PCR on the G3 Radiation Hybrid Panel ${ }^{16}$ ( data not shown); a refinement of the previously published map position ${ }^{17.18}$. A potential translational initiation codon occurs at nt 190, followed by an open reading frame (ORF) of 11,403 bp. A stop codon occurs at nt 11,592 followed by multiple stop codons and a poly(A) tail in a 3 '-untranslated region of 1,933 bp.

Fig. 1 Identification of mutations in three patients with CHS. a, Patient 1, a white male with typical childhood CHS, including oculocutaneous albinism (OCA), neutropenia, impaired platelet function, lack of NK cell activity, and characteristic melanosomal abnormalities 8 Hair and skin are hypopigmented. The patient's parents were third cousins. $b$, A PCR product spanning codons 461-540 exhibited slightly reduced electrophoretic migration in both the SSCP and duplex patterns, suggestive of a small deletion (data not shown). DNA sequence analysis of the PCR product demonstrated that this patient, shown in a, was homozygous for a single-base deletion within codon 489, resulting in a frameshift distal to this site and translational termination at codon 566. c, Sequence analysis of patient 2 revealed a C-T transition resulting in premature termination and presumptive truncation of the polypeptide at codon 1103. Patient 2, a 27-year old white male with late-onset CHS (lymphoblast culture GM03365; Coriell Institute for Medical Research, Camden, NJ), exhibited albinism, recurrent skin infections, neuropathy and mild mental retardation. No additional clinical information or family history is available.


Normal ..... HRH TtR 686 CRH CTT CRT ....

Patient 1 .... R■p Ran Phe lie ......... Pro TIR
566


1183

## Normal

Patient 2
...T\&R
a

b

| 2 | $*$ | Bu b os.i, et ill., 1996 |
| :--- | :--- | :--- |
| CHS <br> patient | $b g 11 \mathrm{~J}$ |  |

p,1tient 3 paticn 11

Fig. 2 a, Modular architecture of CHS and VPS15. Motifs are described in the text. $b$, Schematic representation of the position of known mutations (refs. 14, 15 and presented here) along the length of the CHS protein in both mouse (*) and human ( $\diamond$ )

Comparison of CHS with the 3' partial 7-kb mouse beige cDNA sequence that we reported previously ${ }^{14}$ demonstrates $77.2 \%$ nucleotide identity and $87.9 \%$ amino acid identity between the partial mouse and fulllength CHS sequences. BLAST searches of the GenBank nucleic acid sequence databases identified multiple expressed sequence tags (ESTs), including those previously identified with the murine beige sequence ${ }^{14}$. In addition thirteen human ESTs (N25938, H99579, Z21296, Z21358, N39704, W26957, H50968, M78482, H51623, N74354, W03146, N92032, N74383) exhibited almost complete sequence identity to the human beige cDNA homologue.

The isolation of the human CHS gene allowed us to identify mutations in three patients with CHS (Fig. 1). Patient 1, shown in Fig. la, is an inbred boy with typical childhood CHS $^{8}$ and homozygous for a single-base deletion within codon 489, a frameshift mutation that results in premature translational termination at codon 566 (Fig. lb). Patient 2, an adult male with late-onset CHS, is homozygous for a C to T transition in codon 1103, CGA-to-TGA, resulting in a nonsense mutation (Fig. le). Patient 3, a one year-old girl with typical childhood CHS (fibroblast culture GM02075A; Coriell Institute for Medical Research) exhibits 'partial' OCA, photophobia and cytoplasmic inclusions in her white blood cells. She is heterozygous for a previously described frameshift mutation ${ }^{15}$; a single-base duplication in codon 40, GCA to GGCA (data not shown). We have not yet identified the second mutation in this patient. The overall similarity of the mouse beige and $C H S$ gene sequences and the identification of pathologic mutations in patients with CHS definitively prove that $C H S$ is homologous to mouse beige.

We have recently reported the cloning of the mouse
beige gene ${ }^{14}$ as have Barbosa et al. ${ }^{15}$. However the sequences reported do not align or even partially overlap. Both the partial sequences - that obtained by ourselves and the $4.5-\mathrm{kb}$ sequence obtained by Barbosa et al.- are contained within the larger 11,403bp ORF described here: ours ${ }^{14}$ aligns with the 3 ' end of the ORF of CHS while that of Barbosa et al. aligns with the 5' end. Both groups have also observed a large 12 -kb message in many tissues. The most abundant message observed by Barbosa et al. however, was 7 kb . Based upon its abundance in leukocyte cell lines, they concluded that it corresponds with the message of primary functional significance. Nevertheless, as shown in Fig. 2b, pathologic mutations have been identified along the length of the longer, $12-\mathrm{kb}$ mRNA, suggesting that this longer species is critical to function. Surprisingly, the final 36 amino acids of the murine beige gene reported by Barbosa et al. ${ }^{15}$ are not present in the full-length human CHS polypeptide. However, it is possible to PCR-amplify this segment from both mouse genomic DNA and from a bacterial artificial chromosome contig that completely spans the mouse beige region (M.D.J.et al., manuscript submitted), suggesting that this sequence represents an alternatively spliced exon.

Additional alternative splice forms of CHS have also been isolated; two of the human cDNAs isolated in this study lacked nts 7550-7927. Furthermore, alignment of the murine BG polypeptide sequence reported by Barbosa et al. ${ }^{15}$ with that of the full-length CHS protein revealed the absence of amino acids 1039-1044 in the former. At present it is not clear whether these differences result from the cloning of splice variants, or from true structural differences between the genes in human and mouse.

> 1 MSTDSNSLAR EFLTDVNRLC NAWQRVEAR EEEEEETHHA TLGQYLVHGR
> 51 GF LLLTKLNS IIDQALTCRE ELLTLL LSLL PLVWKI P VQE EKATDFNLPL 101 s,orrLTKEK NSSSQRSTQE KLHLEGS,LS SQVS"'VNVF RKSRRQRKIT 151 HR YSV RDARK TQLSTSDSEA NSOEKGIAMN KHRRPHLLHH FLTSFPKQDH 201 PKAKLDRLAT KEQTPPDAMA LENSREIIPR QGSNTDILSE PAALSVISNM 251 NNSPFDLCHV LLSLLEKVCK FDVTLNHNSP LAASWPTLT EFLAGFGDCC 301 SLSDNLESRV VSAGWTEEPV ALIQRMLFRT VLHLLSVDVS TAEMMPENLR 3 51 KNLTELLRAA LKIRICLEKQ PDPFAPRQKK TLQEVQEDFV FSKYRH RALL 401 LPELLEGVLQ ILICCLQSAA SNPFYFSQAM DLVQEFIOHH GP NLF ETAVL 451 QMEWLVLR[X; VPPEASEHLK ALINSVMKIM STVKKVKSEQ LHHSMCTRKR 501 HRRCEYSHFM HHHRDLSGLL VSAFKNQVSK NPFEETAOOD VYYPE RCCCI 551 AVCAHQCLRL LQQASLSSTC VQI LSGVHNI GI CCCMDPKS VII P LLHAFK 601 LP ALKNFOOH I LNI LNKLI L DQLGGAEISP KI KKAACNI C TVDSDQLAQL 651 EETLQGNLCD AELSSSLSSP SYRFQGILPS SGSEDLLWI<W DALKAYQNFV 701 FEEDRLHSIQ IANHICNLIQ KGNIVVQWKL YNYI FN PVLQ RGVELAHHCQ 751 HLSVTSAQSH VCSHHNQCLP QDVLQIYVK T LPILLKSRVI RDL FLSCNGV 801 SQIIELNCLN GIRSHSLKAF ETLIISLGEQ QKDASVPDIO GIDIEQKELS 851 SVJNG TSFHH QQAYSDS PQS LSKFYAGLKE AYPKRRKTVN QOVHI NTINL 901 FLCVAF LCVS KEAESDRESA NDSEDTSGYD ST AS EPLS HM LPCISLES LV 951 LPSPE HMHQA ADIWSMCRWI YMLSSVFQKQ FY RLGGFRVC HKLI FMI IQK 1001 LFRSHK EEQG KKEGDTSVNE NQDLNRISQP KRTHKEOLLS LAIKSDPIPS 1051 ELGS LKKSA D SLGKLELQHI SSINVEEVSA TEAAPEEAK L FTSQESETS L 1101 QSIRLLEALL AICLHGARTS QQKMELELP N QNLSVE SI LF EMRDHLSQSK 1151 VI ETQLAKPL FDALLRVALG NYSADFEHND AMTEKSHQSA EELSSQ PGDF 1201 SEEAEDSQCC SFKLLVEEEG YEADSESNPE DGETQDDGVD LKSETEGFSA 1251 SSSPNDLLEN LTQGEIIYPE ICMLELNLLS ASKAKLDVLA HVF ESFLKII 1301 RQKEKNVFL L MQQGTVKNLL GGPLSJLTQD DSDF QACQRV LVDLLVSLMS 1351 SR T CS EELTL LLR I F LEKSP C TKI LLLGI L KII ESOTTMS PS QYLTFPLL 14 01 HAPNLSNGVS SQKYPGILNS KAMGLLRRAR VSRSKKEAOR E.SFPHR LLSS 1451 WHI APVHLPL LGQNCWPH LS EGFSVSL WFN VECIHEAEST TEKGKKIKKR 1501 NKS LI LPDSS FDGTESDRPE GAEYINPGER LIEEGCIHII SLGS KAL MI Q 1551 VWADPHNATL IFRVCMDSNO OMKAVLLAQV ESQENI F LPS KWQHLVLTYL 1 601 QQPQGKRRIH GKISI WVSG Q RKPDVTLDFM LPR KTSLSSD SNKTFCMIGH 16 51 CLSSQEEFLQ LAGKWDLGNL LLFNGAKVGS QEAFYLYACG PNHTSVMPCK 170 1 YGKP VNOYSK Y I NK R.I LRCE QI RELF MTKK DVDIGLL! ES LSWYTTYCP 1751 AQYTI YEPVI RLKGQMK TQL SQRPFSS KEV QSI LLEPHHL KNLQPTEYKT 1801 I QGI LHEI CG TGI FVF LFAR VVELSSCEET QALALRVILS LIKYNQQRVH 18SI ELENC NGLSM IHQVLIKQKC IVGFY!LKTL LEGCCGED1 I YMNENGEFKL 1901 DVDSNAIIQD VKLLEELLLD WK I WS KAEQG VWETLL AAL I:: VLI RADHHQQ 1951 MFN I KQLLKA QW HHF LLTC QVLQF.YK EGQ LTP MPR EVCR S FVK II AEVL 200 GSPPDL ELLT IIFNFLLAVH PPTNTYVCHN PTNFYFSLHI DGKIFQEKVR 2051 SI MYLRH SSS GGRS LMS PGf MVISPSGFTA SPYEGENSSN IIPQQMAAHM 2101 LRSRSLPAFP TSSL LTQSQK LTGS LGCS I D R[,QNI ADTYV ATQSKKQNS L 2151 GSSDTLKKGK EDAF1SS CES AKTVCEHEAV LSAQVSVS OV PKGVLGFPW 2201 KADHKQLGAE PRSEDDSP GD ESCPRRPD YL KGLASFQRSH STIASLGLAF 2251 PSQNGSMVG RWPSLVDRNT DDWENFAYSL GYEPNYNRTA SAHSVTEOCL 2301 VPI CCGLYEL LSGVLLILPD VLLED VMD KL I QADTLL VLV NHPSPAIQQG 2351 VIKLLDAYFA RASKEQKDKF LKNRGFSLLA NQLYLHR GTQ ELLECFIEMF 2401 FGRHIGLDEE FDLEDVRNMG LFQKWSVIPI LGLIETSLYD NILLHNALLL 2451 LLQI LNSCSK VADMLLDNGL LYVLCNTVAA LNGLEKNI PM S EYKLLACDI 2501 QQLFI AVTI H ACSSSGSQYf RVI EDLI VML GYLQNSKNKR TQNMAVALQL 2551 RV LQAAMEFI RTTANHD S EN LTDSLQS PSA PHHAVVQKRK S IA GPRKFP L 2601 AQTESLLMKM RSVANDELHV MMQRRMSQE.N PS QATETELA QRLQRLTVLA 265 VNR II YQ8 FN SDIIDILRTP ENVTQSKTSV FQTEISEENI HHEQSSVFNP 2701 FQKE I FTYLV EGFKVSIGSS KASGS KQQWT KILWSCKETF RMQLGRLLVH 2751 I LS PAHAAQE RKQIFEIVHE PNHQEILRDC LSPSLQHGAK LVLYLSELIH 2801 NHQGEL TEEE LGTAELLMNA LKLCGHKCI P PSA S TKADLI KMIKEEQKKY 2851 E'TEEGVNKAA WQKTVNNNQO SLFQRLDSKS KDISKIAADI TQAVSLSQGN 2901 ERKKVIQHI R GMYKVDLSAS RHWQELIQQL THDRAVWYDP IYYPTSWQLD2951 PTEGPNRERR RL QRCYLTI P NKYLLRDRQK SEDVVKPPLS YLFEOKT HSS 3001 FSSTVKOKAA SES IR VNRRC IS VA PS RET A GEl LLGKCGM 'ffVEDNASDT 3051 VESSSLQGEL EPASFSWTYE EI KEVHKR WW QLRDNAVEI F LTNGRTLLLA 3101 FDNTKVRODV YHNILTNNLP NL LEYGNI TA LTNLWYTGQI TNF EYL THLN 31SI KHAGRSf NDL MQYPVFPF IL ADYVSETLDL NDLLI YRN LS KPIAV QYKEK 3201 EDRYVO TYKY LEEEYRKGAR EDDPMPPVQP YHYG SHYSNS GTVLHFLVRM 3251 PPFTKMfLAY QOOSFDIPDR TFHS TNTIWR LSSFE SMT OV KELIPEFFYL 330 PEFLVNREGF DFGVRQNGER VN HVNLPPWA RNDPRLF I LI HRQALESDYV 3351 SQ NI CQWI DL VFGYKQKGKA SVQA itN FHP ATYF MDVSA VEDPV'QRRAL 3401 ETMI KTYG QT PRQLFHMAHV SR PGAK LNI E GEL PAAVGLL VQFA FRETRE 3451 QVKEITYPSP LSWIKGLKWG EYVGSPSAPV PW CFSQPH G ERFGSLQALP 3501 TRAI CGLSRN FCLVMTYSK E QGVRS MNS TD I QWS AI LSWG YADN I LRLKS 3551 KQ S EPP VNFI QSS QQYQVTS CAWVPDS CQL FTGSKC GVI T AYTNRF TSST 3601 PSEI EMETQI HLYGHTEEIT SLF VCK PYSI LI SVSR DGTC II WDLNRLC Y 3651 VQSLAGHKSP VTAVSASET S GDIATVCDSA GGGSDLRLWT VNGDLVGHVH 3701 CRE II C'SVA F S NQPEGVSI N VI AGGLENGI VRLWS TWDLK P VREIT FP KS 3751 NKPII SLTFS CDGHHLYTAN SDGTVIAWCR KDQORL KQPM FYSFLSSYAA 380 1G

> Fig. 3 Amino acid sequence of the CHS gene product.


Fig. 4 Alignment of the CHS protein sequence and related protein sequences: Saccharomyces cerevisiae ORF, YCR032w (Genbank \#P25356), Caenorhabditis elegans ORFs T01H10.8 (Genbank \#1054706), F10F2.1 (Genbank accession number 1066956) YSM3_CAEEL (SwissProt \#010123), YSM2_CAEEL (SwissProt accession number 010122) and human cell division control protein 4-related protein (CDC4L) (Genbank \#A43289). The alignment was performed as described ${ }^{14}$. Residues highlighted in yellow indicate identities; residues highlighted in blue indicate conservative substitutions

The 11,403-bp open reading frame (ORF) of CHS predicts a polypeptide of3,80l amino acids (Fig. 3) with a molecular mass of 429,153 . BLAST searches of the NCBI and SwissProt protein databases reveal significant homology between CHS and related proteins pre-
viously reported for murine BG ${ }^{14}$. These proteins include a Saccharomyces cerevisiae ORF, a human cell division control protein 4-related protein (CDC4L) and two anonymous Caenorhabditis elegans ORFs. In addition, a third homologue was identified in C. elegans by
merging YSM3_CAEEL and YSM2_CAEEL into one contiguous sequence that appears to contain two genes both related to BG/CHS (Fig. 4). Both significantly match adjacent regions in the portion of CHS that is conserved, and represent two separate loci that are physically linked within the C. elegans genome. Of all the ORFs described above, F10F2.1 in C.elegans most closely resembles those of $b g$ and $C H S$. As much of this sequence homology is restricted to amino acids 3116-3461, we conclude that this is a functionally and structurally defined domain. We have designated this conserved domain BEACH (BEige And CHS) because there is high conservation in otherwise distinct proteins over a wide species range, the length of the region is much larger than protein-protein interaction domains, and because of similarities in predicted alpha-beta folding type and the clustering and properties of the conserved amino acids ${ }^{14}$

The human CHS polypeptide also contains a number of sequence motifs that may provide further clues to its biological function (Fig. 2a). The CHS protein consists of a series of hydrophobic helices, with interspersed hydrophobic regions that do not appear to represent transmembrane domains but are consistent with transmembrane association. These helical regions most closely resemble ARM ${ }^{20}$ and $\mathrm{HEAT}^{21}$ repeat motifs, which tend to form long rods ${ }^{20,21}$. HEAT repeat motifs occur as a minimum of three repeats in tandem in extremely large proteins that contain extensive helical regions, not unlike CHS. Many of the known HEAT repeat motif-containing proteins are associated with vesicle transport ${ }^{22-25}$. The C-terminal region of the CHS polypeptide contains seven consecutive WD40 motifs. Four of these precisely correspond with the WD40 consensus, and three have minor deviations ${ }^{26}$. Such consecutive WD40 motifs form beta sheets arranged in a 7-bladed beta 'propeller fold ${ }^{27,28}$ that is thought to mediate protein-protein interactions.

Barbosa et al. ${ }^{15}$ reported that the mouse $B G$ protein exhibits homology to stathmin, a phosphoprotein that regulates the polymerization of microtubules ${ }^{29}$. We also observe a $26 \%$ identity between $\mathrm{BG} / \mathrm{CHS}$ and stathmin across 72 amino acids (464-536). However, the function of stathmin is dependent on this region forming a coiledcoil, but our analysis of $\mathrm{BG} / \mathrm{CHS}$ across this region indicates that it has no coiled-coil potential as calculated by using a coiled-coil prediction program ${ }^{30}$. Furthermore, this small region merely represents $2 \%$ of the total size of the protein and comparison of the full-length human CHS polypeptide further weakens the significance of this alignment ( $1.3 \times 10^{-4}$ probability by chance). Thus, the significance of the 'homology' between BG/CHS and stathmin remains questionable.

As noted above, the CHS polypeptide shares the closest homology to a number of anonymous ORFs. In addition, the modular architecture of sequence motifs in CHS suggests some potential functions. The presence of multiple hydrophobic regions resembling related ARM and HEAT repeats ${ }^{19}$, as well as $4-7$ consecutive WD40 sequence motifs, suggests that CHS may be a cytoplasmic protein involved in transport and/or associated with vesicles. The only known protein that contains HEAT repeats (or helical regions that resemble HEAT and ARM repeats), C-terminal consecutive WD40 motifs and a globular alpha/beta domain is the yeast serine/threonine
protein kinase, VPS15. Vps15 is a member of a large class of yeast ' $V p s$ ' mutants that are associated with defective vacuolar protein sorting, and fall into more than 40 complementation groups (reviewed in ref. 31). Vps mutants exhibit defective sorting in that they tend to secrete soluble vacuolar hydrolase precursors instead of sorting them into the vacuolar compartment ${ }^{32}$. It is thought that VPS15 is required for activation of a second VPS (VPS34) and that they function together as components of a membrane-associated signal transduction complex that regulates intracellular protein trafficking ${ }^{33}$.

Given the similarity in modular architecture of VPS15 and the BG/CHS proteins (Fig. 2a), we suggest that the BG/CHS proteins may have a similar function, consistent with the observation of defective vesicular transport to and from the lysosome and late endosome ${ }^{6,8,10,12}$ and aberrant compartmentalization of lysosomal and granular enzymes ${ }^{5,7,8,10,11}$ both in humans with CHS and in mice carrying beige mutations. The role of BG/CHS as a component of a membrane-associated signal transduction complex that regulates intracellular protein trafficking is also consistent with our observations that multiple BG/CHS paralogues exist in C. elegans. This suggests that $\mathrm{BG} / \mathrm{CHS}$ may define a novel gene family; an hypothesis compatible with both the existence of multiple mouse loci, mutants of which have phenotypes that are partially similar to beige mutants ${ }^{34}$, and with the potential of CHS being a heterogeneous disease ${ }^{17}$.

## Methods

CHS CDNA isolation. To isolate the complete human beige cDNA homologue, we first screened human fetal liver and retina cDNA libraries (Clontech) with probes from multiple regions of the murine beige cDNA ${ }^{14}$. Subsequently, we carried out multiple rounds of screening using probes derived from the newly isolated human beige homologous cDNAs. In total, 27 cDNAs ( 12 from a fetal liver library and 15 from a human retina library) were identified and sequenced completely on both strands.

Genetic mapping. Sequence-tagged sites (STSs) were designed to multiple segments along the length of the CDNA; all STSs were mapped by PCR on the G3 Radiation Hybrid Panel ${ }^{16}$.

Mutation detection. High molecular weight genomic DNA was prepared from cultured cells from patients with CHS by standard methods. DNA segments were amplified from genomic DNAs of the patients and at least two normal individuals by PCR using primers derived from the cDNA sequence, and analysed by simultaneous single-strand conformation polymorphism (SSCP)/heteroduplex (HDX) analyses. Segments exhibiting aberrant SSCP/HDX patterns were either reamplified from each patient in duplicate, and cloned into pCR2.1 (InVitrogen, San Diego, CA) from which at least six independent clones were sequenced, or sequenced directly from multiple independent PCR amplifications.

Structural analysis. Initial database searches were performed using the BLAST series of programs ${ }^{35}$. Putative nonglobular domains that function as linkers were identified using the SEG program (ref. 36 and refs therein). Secondary structure prediction and screening for transmembrane helices were carried out using the PHD Web server implement ${ }^{37}$. The coiled-coil potential was measured using the COILS program ${ }^{30}$. To increase the sensitivity of the database similarity search programs, CHS was partioned into segments according to similar secondary struc-
ture content and to putative globular domains. As homologous sequences were identified, iterate motif and profile searches were performed ${ }^{38}$.

GenBank accession number. The accession number for the CHS cDNA sequence is U67615

1. Beguez-Cesar, A.B. Neutropenia cronica maligna familiar congranulaciones atipicas de los leucocitos. Bol. Soc. Cubana Pediatr. 15, 900-922 (1943).
2. Steinbrinck, W. Uber eine neue Granulationsanomalie der Leukocyten. Dtsch. Arch. Klin. Med. 193, 577-581 (1948).
3. Chediak, M. Nouvelle anomalie leukocytaire de caractere constitutionnel et familiel. Rev. Hematof. 7. 362-367 (1952).
4. Higashi, O. Congenital gigantism of peroxidase granules. Tohoku J. Exp. Med. 59, 315-332 (1954)
5. Jones, K.L., Stewart, R.M., Fowler, M., Fukuda, M. \& Holcombe, R.F. Chediak-Higashi lymphoblastoid cell lines: granule charactenstics and expression of lysosome-associated membrane proteins. Cfin. immunol. Immunopath. 65, 219-226 (1992)
6. Burkhardt, J.K., Wiebel., F.A., Hester, S. \& Argon, Y. The giant organelles in Beige and Chediak-Higashi fibroblasts are derived from late endosomes and mature lysosomes. J. Exp. Med. 178, 1845-1856 (1993).
7. Holcombe, R.F., Jones, K.L. \& Stewart, R.M. Lysosomal enzyme activities in Chediak-Higashi syndrome: evaluation of lymphoblastoid cell lines and review of the literature. Immunodeficiency 5, 131-140 (1994).
8. Zhao, H. et al. On the analysis of the pathophysiology of Chediak-Higashi syndrome. Lab. Investig. 71, 25-34 (1994).
9. Lutzner, M.A., Lowrie, C.T. \& Jordan, H.W. Giant granules in leukocytes of the beige mouse. Heredity 58, 299 -300 (1966)
10. Brandt, E.J., Elliott, R.W. \& Swank, R.T. Defective lysosomal enzyme secretion in kidneys of Chediak-Higashi (beige) mice. J. Cell Biol. 67, 774-788 (1975).
11. Swank, R.T. \& Brandt, E.J. Turnover of kidney $\beta$-glucuronidase in normal and Chediak-Higashi (beige) mice. Am. J. Pathol. 92, 755-771 (1978).
12. Willingham, M.C., Spicer, S.S. \& Vincent, R.A., Jr. The origin and fate of large dense bodies in beige mouse fibroblasts. Exp. Cell Res. 136, 157-168 (1981).
13. Penner, J.D. \& Prieur, D.J. A comparative study of the lesions in cultured fibroblasts of humans and four species of animals with Chediak-Higashi syndrome. Am. J. Med. Genet. 28, 445-454 (1987).
14. Perou, C.M. et al. Identification of the murne beige gene by YAC complementation and positional cloning. Nature Genet. 13, 303-308 (1996).
15. Barbosa, M.D.F.S. et al. Identification of the homologous beige and Chediak-Higashi syndrome genes. Nature 382, 262-265 (1996)
16. Stein, L., Kruglyak, L., Slonim, D. \& Lander, E. Unpublished software, Whitehead Institute/MIT Center for Genome Research (1995).
17. Fukai, K. et al. Homozygosity mapping of the gene for Chediak-Higashi syndrome to chromosome 1q42-q44 in a segment of conserved synteny that includes the mouse beige locus (bg). Amn. J. Hum. Genet., 59, 620-624 (1996).
18. Barret, F.J. et al. Genetic and physical mapping of the Chediak-Higashi syndrome on chromosome 1q42-43. Am. J. Hum. Genet. 59, 625-632 (1996).
19. Bork, P. \& Koonin, E.V. Protein sequence motifs. Curr. Opin. Struct. Biol. 6, 366-376 (1996).

## Acknowledgements

We thank N. Lakey and members of the Sequencing group at MPI, and J. Kaplan for helpful discussions.

## Received 4 September; accepted 19 September 1996.

20. Peifer, M., Berg, S. \& Reynolds, A.B. A repeating amino acid motif shared by proteins with diverse cellular roles. Cell 76, 789-791 (1994).
21. Andrade, M.A. \& Bork, P. HEAT repeats in the Huntington's disease protein. Nature Genet. 11, 115-116 (1995).
22. DiFiglia, M. et al. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. Neuron 14, 1075-1081 (1995).
23. Sabatini, D.M., Erdjument-Bromage, H., Lui, M., Tempst, P. \& Snyder, S.H RAFT1: a mammalian protein that binds to FKBP12 in a rapamycindependent fashion and is homologous to yeast TORs. Cell 78, 35-43 (1994).
24. Zheng, X. F., Florentino, D., Chen, J., Crabtree, G. R. \& Schreiber, S. L. TOR kinase domains are required for two distinct functions, only one of which is inhibited by rapamycin. Cell 82, 121-130 (1995).
25. Hemmings, B. A. et al. $\alpha$ - and $\beta$-forms of the $65-\mathrm{kDa}$ subunit of protein phosphatase 2A have a similar 39 amino acid repeating structure. Biochemistry 29, 3166-3173 (1990).
26. Neer, E.J., Schmidt, C.J., Nambudripad, R. \& Smith, T.F. The ancient regulatory-protein family of WD-repeat proteins. Nature 371, 297-300 (1994).
27. Sondek, J., Bohm, A., Lambright, D. G.. Hamm, H. E. \& Sigler, P. B. Crystal structure of a GA protein beta gamma dimer at 2.1 A resolution. Nature 379 , 369-374 (1996).
28. Wail, M. A. et ad. The structure of the $G$ protein heterotrimer $\mathrm{G}^{\prime} \alpha 1 \beta 1 \mathrm{Y} 2.83$ 1047-1058 (1995)
29. Belmont, L. D. \& Mitchison, T. J. Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. Cell 84 623-631 (1996).
30. Lupas, A.N. et a!. Predicting coiled coils from protein sequences. Science 252, 1162-1164 (1991)
31. Klionsky, D. J. \& Emr, S. D. A new class of lysosomal/vacuolar protein sorting signals. J. Biol. Chem. 265, 5349-5352 (1990).
32. Herman, P. K., Stack. J. H. \& Emr, S. D. A genetic and structural analysis of the yeast Vps 15 protein kinase: evidence for a direct role of VPS $15 p$ in vacuolar protein delivery. EMBO J. 10, 4049-60 (1991).
33. Stack, J. H., Herman, P. K., Schu, P. V. \& Emr, S. D. A membrane-associated complex containing the Vps15 protein kinase and the VPS34 PI 3-kinase is essential for protein sorting to the yeast lysosome-like vacuole. EMBO J. 12, 2195-204 (1993).
34. Novak, E.K., Hui, S.W. \& Swank, R.T. Platelet storage pool deficiency in mouse pigment mutations associated with seven distinct genetic loci. Blood 63, 536-544 (1984).
35. Altschul, S. F., Boguski, M. S., Gish, W. \& Wootton, J. C. Issues in searching molecular sequence databases. Nature Genet. 6, 119-129 (1994).
36. Wootton, J. C. \& Federten. S. Analysis of compositionaily biased regions in sequence databases. Meth. Enz. 266, 554-571 (1996).
37. Rost, B., Sander, C. \& Schneider, R. PHD-an automatic mail server for protein secondary structure prediction. Comput. Appl. Biosci. 10, 53-60 (1994).
38. Bork, P. \& Gibson, T. Applying motif and profile searches. Meth. Enz. 266, 162-184 (1996).
