15. E. Sato, R. A. Furuta, T. Miyazawa, *Retrovirology* **7**, 110 (2010).
16. K. Fukuda *et al.*, *Ann. Intern. Med.* **121**, 953 (1994).
17. Materials and methods are available as supporting material on *Science* Online.
18. G. Simmons *et al.*, *Transfusion* **51**, 643 (2011).
19. O. E. Varnier, A. D. Hoffman, B. A. Nexø, J. A. Levy, *Virology* **132**, 79 (1984).
20. O. E. Varnier, C. M. Repetto, S. P. Raffanti, A. Alama, J. A. Levy, *J. Gen. Virol.* **64**, 425 (1983).
21. A. D. Hoffman, B. Banapour, J. A. Levy, *Virology* **147**, 326 (1985).
22. X. Qiu *et al.*, *Retrovirology* **7**, 68 (2010).
23. B. Banapour, J. Sernatinger, J. A. Levy, *Virology* **152**, 268 (1986).
24. Y. Takeuchi *et al.*, *J. Virol.* **68**, 8001 (1994).
25. R. P. Rother *et al.*, *J. Exp. Med.* **182**, 1345 (1995).
26. D. M. Takefman, G. T. Spear, M. Saifuddin, C. A. Wilson, *J. Virol.* **76**, 1999 (2002).
27. T. Paprotka *et al.*, *J. Virol.* **84**, 5719 (2010).
28. H. C. Groom, M. W. Yap, R. P. Galão, S. J. Neil, K. N. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5166 (2010).
29. R. A. Weiss, *BMC Biol.* **8**, 124 (2010).
30. S. Hué *et al.*, *Retrovirology* **7**, 111 (2010).
31. A. L. Landay, E. T. Lennette, C. Jessop, J. A. Levy, *Lancet* **338**, 707 (1991).

Acknowledgments: We thank D. Peterson for the referrals of all the subjects evaluated, J. Weismann for coordinating subject participation, I. Livinti for assistance with data management, E. Delwart and J. Fiddes for comments on the manuscript, and K. Peter for help in preparation of the manuscript. These studies were conducted with support of private funds to the investigators. Patent applications have been filed by Abbott Laboratories relating to detection of XMRV using immunoassays and molecular-based assays.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1204963/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S3
References

1 March 2011; accepted 16 May 2011
Published online 31 May 2011;
10.1126/science.1204963

Recombinant Origin of the Retrovirus XMRV

Tobias Paprotka,[1]* Krista A. Delviks-Frankenberry,[1]* Oya Cingöz,[3,4]* Anthony Martinez,[5] Hsing-Jien Kung,[5,6] Clifford G. Tepper,[5] Wei-Shau Hu,[2] Matthew J. Fivash Jr.,[7] John M. Coffin,[3,4] Vinay K. Pathak[1]†

The retrovirus XMRV (xenotropic murine leukemia virus–related virus) has been detected in human prostate tumors and in blood samples from patients with chronic fatigue syndrome, but these findings have not been replicated. We hypothesized that an understanding of when and how XMRV first arose might help explain the discrepant results. We studied human prostate cancer cell lines CWR22R*v1* and CWR-R1, which produce XMRV virtually identical to the viruses recently found in patient samples, as well as their progenitor human prostate tumor xenograft (CWR22) that had been passaged in mice. We detected XMRV infection in the two cell lines and in the later passage xenografts, but not in the early passages. In particular, we found that the host mice contained two proviruses, PreXMRV-1 and PreXMRV-2, which share 99.92% identity with XMRV over >3.2-kilobase stretches of their genomes. We conclude that XMRV was not present in the original CWR22 tumor but was generated by recombination of two proviruses during tumor passaging in mice. The probability that an identical recombinant was generated independently is negligible (~10^{-12}); our results suggest that the association of XMRV with human disease is due to contamination of human samples with virus originating from this recombination event.

Murine leukemia viruses (MLVs) are retroviruses belonging to the genus *Gammaretrovirus* that cause cancers and other diseases in mice, and they are divided into the ecotropic, amphotropic, polytropic, and xenotropic classes on the basis of their receptor usage. Xenotropic MLVs cannot infect cells from inbred mice but can infect cells from other species, including humans. Xenotropic murine leukemia virus–related virus (XMRV) was isolated from a human prostate cancer (PC) in 2006 and has been reported to be present in 6 to 27% of human PCs (*1*, *2*) and in the peripheral blood of 67% of chronic fatigue syndrome (CFS) patients (*3*). The assertion that XMRV is circulating in the human population has been challenged by several studies that have failed to detect XMRV in multiple cohorts of PC and CFS patients or healthy controls [reviewed in (*4*)]. Endogenous xenotropic MLVs can infect human tumors during passage through nude mice (*5*), and it has been suggested that XMRV may have arisen in this manner (*5*, *6*). In addition, XMRV replication is highly sensitive to human APOBEC3s and tetherin (*7*–*11*), which makes it doubtful that XMRV replication occurred efficiently in human peripheral blood mononuclear cells of CFS patients as previously reported (*3*).

The human PC cell line CWR22R*v1* (hereafter 22R*v1*) (*12*) produces infectious XMRV essentially identical in sequence to that obtained from patients. 22R*v1* contains ≥10 proviral copies per cell (*13*) and was proposed to have been derived from an XMRV-infected tumor. This cell line was derived from a xenograft (CWR22) that was established from a primary prostate tumor at Case Western Reserve University and serially passaged in nude mice (*14*, *15*). To explore the origin of the virus in 22R*v1* cells, we analyzed various passages of the CWR22 xenograft, as well as a subline of the CWR22 xenograft (2152) from which the 22R*v1* cell line was established (*12*), and another prostate cancer cell line, CWR-R1, which was also derived from CWR22 (*16*). Figure 1A traces the timeline of the serial xenograft transplants of CWR22 up to the derivation of the cell lines 22R*v1* and CWR-R1 and indicates (bold letters) the samples that were available for analysis. Nude mouse strain(s) maintained by Charles River (NU/NU) and Harlan Laboratories [Harlan Sprague Dawley (Hsd)] are likely to have been used for in vivo passages of the xenograft (*17*). DNA samples from passage 3 (777 in Fig. 1A) and an unknown early passage (736) were obtained along with samples from a 7th passage, CWR22-9216R and CWR22-9218R. A xenograft tumor from the early seventh passage was independently propagated at the University of California, Davis, using Hsd nude mice (CWR22-8R and 8L). Total nucleic acid from relapsed androgen-independent tumors (CWR22R) 2152, 2524, 2272, and 2274 and the 22R*v1* and CWR-R1 cell lines was available for analysis (*14*).

We verified that the xenograft samples (736, 777, 9216R, 9218R, 8R, and 8L) and the 22R*v1* or CWR-R1 cell lines were all derived from the same person by performing short tandem repeat analysis at seven loci (Fig. 1B and fig. S1). The probabilities that the xenografts and the two cell lines have the same allele patterns for these loci by chance are 1.6×10^{-13} and 6.3×10^{-13}, respectively.

To quantify the amount of XMRV DNA in the CWR22 xenografts, we developed a real-time polymerase chain reaction (PCR) primer-probe set that specifically detected XMRV *env* and excluded murine endogenous proviruses present in BALB/c and NIH3T3 genomic DNA (Fig. 1C). We used quantitative PCR of 22R*v1* DNA to estimate 20 proviruses per cell and used the 22R*v1*

[1]Viral Mutation Section, HIV Drug Resistance Program, National Cancer Institute at Frederick, Frederick, MD 21702, USA. [2]Viral Recombination Section, HIV Drug Resistance Program, National Cancer Institute at Frederick, Frederick, MD 21702, USA. [3]Department of Molecular Biology and Microbiology, School of Medicine, Tufts University, 150 Harrison Avenue, Boston, MA 02111, USA. [4]Genetics Program, School of Medicine, Tufts University, 150 Harrison Avenue, Boston, MA 02111, USA. [5]Department of Biochemistry and Molecular Medicine, University of California, Davis, Sacramento, CA 95817, USA. [6]Department of Urology, University of California, Davis, Sacramento, CA 95817, USA. [7]Data Management Services, Inc., National Cancer Institute at Frederick, Frederick, MD 21702, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: vinay.pathak@nih.gov

DNA to generate a standard curve. The CWR22 xenografts had significantly fewer copies of XMRV *env* (<1 to 3 copies per 100 cells) compared with the 22R*v1* cells (2000 copies per 100 cells). The CWR-R1 cell line had ~3000 copies per 100 cells, and the NU/NU and Hsd nude mice, thought to have been used to passage the CWR22 xenograft, had 58 and 68 copies per 100 cells, respectively. Because xenograft tumors are expected to contain a mixture of human and mouse cells, we quantified the amount of mouse DNA by analyzing mouse intracisternal A-type particle (IAP) DNA as previously described (*18*, *19*). About 0.3 to 1% of the total DNA from all six xenografts consisted of mouse DNA (Fig. 1D); this result is consistent with the <1 to 3 XMRV *env* sequences per 100 cells detected in the same samples (Fig. 1C).

We characterized XMRV and related sequences in the xenografts, cell lines, and nude mouse strains by PCR and DNA sequencing (Fig. 2). Using primers previously used to clone and sequence XMRV from 22R*v1* cells (*8*), we determined that all the XMRV proviruses in the CWR-R1 and 22R*v1* cell lines are identical in sequence (GenBank accession no. FN692043), with the exception of some rare hypermutated proviruses (Fig. 2A and figs. S2 and S3). Next, we developed several primer sets specifically to amplify XMRV sequences and to exclude endogenous murine retroviruses (fig. S2). Primers that specifically amplified XMRV were used to perform PCR on DNA from the late-passage xenografts 2152, 2524, 2272, and 2274; sequencing confirmed the presence of these XMRV sequences in these tumors (Fig. 2A and fig. S3A; boxed in Fig. 1A).

We used the same XMRV-specific primer sets to amplify and sequence DNA from early-passage xenografts (736, 777, 8L, 8R, 16R, and 18R) (Fig. 2B); the results showed that XMRV *env* was present but not *gag* sequences (sequencing coverage summarized in fig. S3), which indicated that the early xenografts did not contain XMRV. However, we did find that early xenografts contained a previously undescribed XMRV-related provirus that we have named PreXMRV-1 (Fig. 2B). The complete sequence of PreXMRV-1 was determined from the early-passage xenografts, the NU/NU and Hsd strains, and the CWR-R1 cell line (GenBank accession no. FR871849). PreXMRV-1 and consensus XMRV differed by only one base in a 3211-nucleotide (nt) stretch of the genome encoding the 3′ half of *pol* and the 5′ two-thirds of *env*. In addition, the long terminal repeats (LTRs) were nearly identical; PreXMRV-1 had a single adenine deletion relative to XMRV in a run of six adenines. The two genomes differed by 10% over the remaining 3.5-kb stretch of *gag-pro-pol* and by 9% in a 600-nt stretch at the 3′ end of *env*. PreXMRV-1 is replication-defective because of a 16-nt deletion in *gag* and a +1 frameshift mutation in *pol*. Late-passage xenografts 2524 and 2274, but not 2152 and 2272, also contained PreXMRV-1. The detection of low levels of XMRV *env* sequence in the early xenografts (Fig. 1C) can be attributed to the PreXMRV-1 proviruses present in the contaminating mouse DNA. Overall, these results indicate that PreXMRV-1 is an endogenous murine provirus that is present in the NU/NU and Hsd strains, but neither of these strains contains XMRV

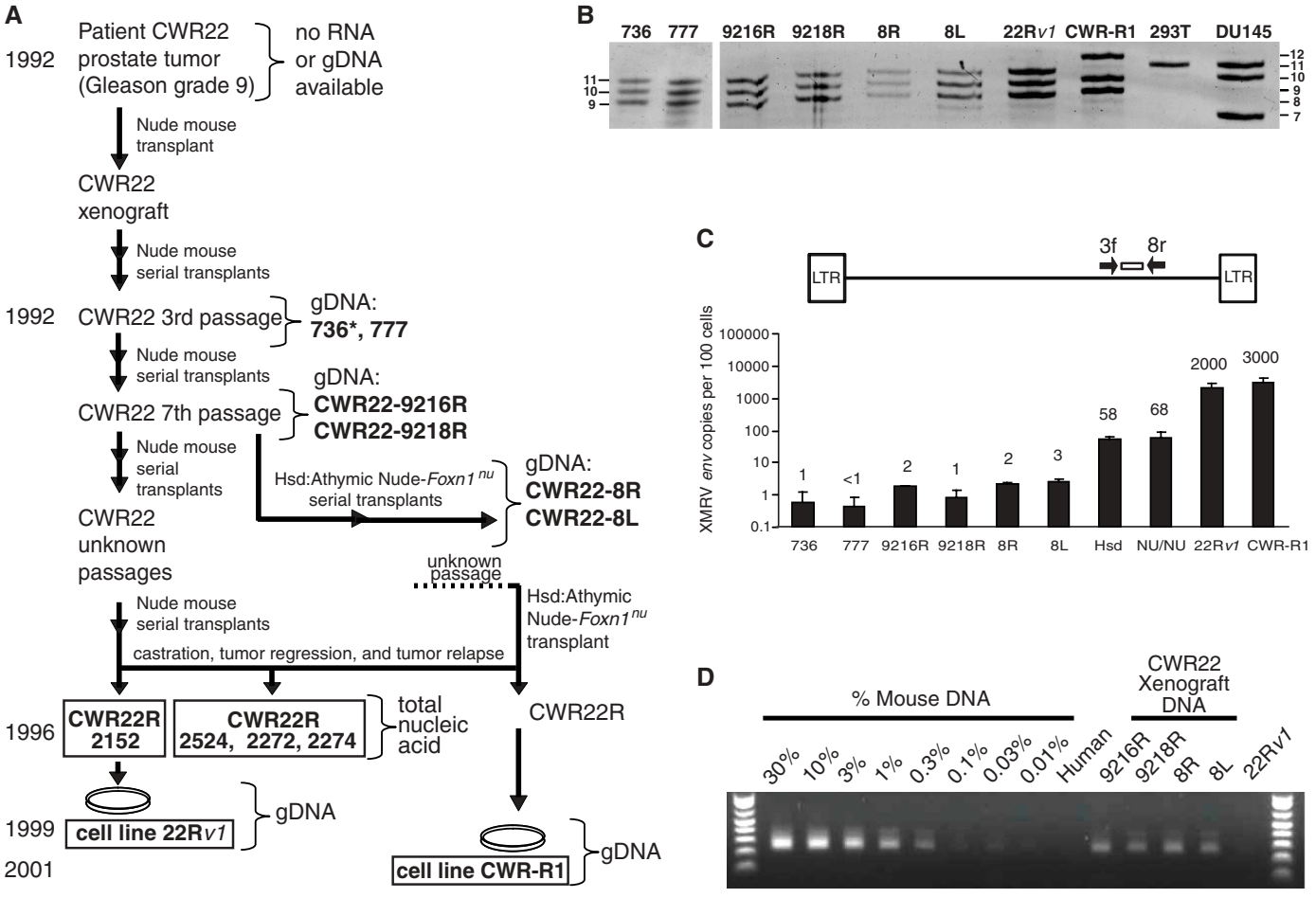


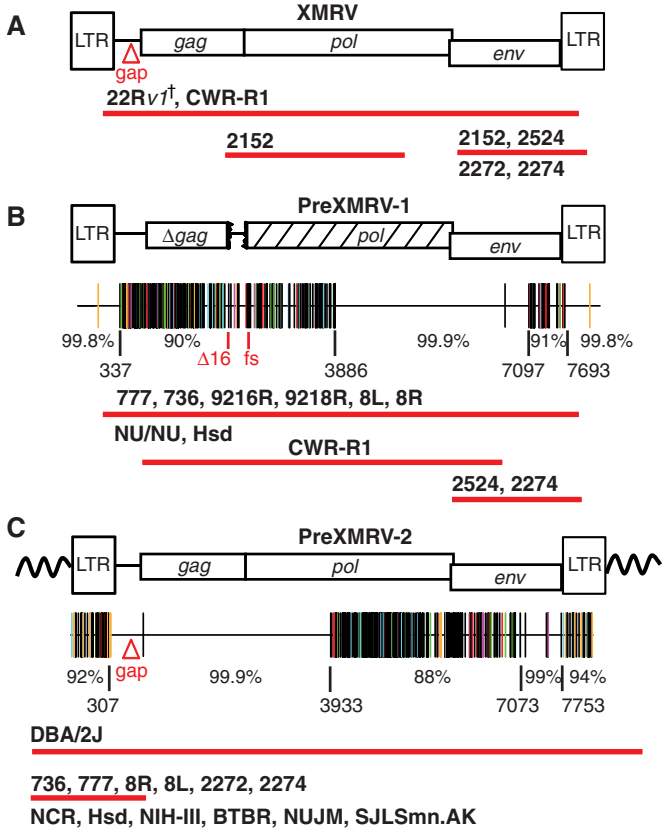
Fig. 1. Characterization of CWR22 xenografts and XMRV-related sequences. (**A**) Genesis of 22R*v1* and CWR-R1 cell lines. Bold letters indicate samples from which genomic DNA (gDNA) or total nucleic acid was available for analysis. XMRV-positive samples are boxed. Asterisk (*) indicates unknown early passage. (**B**) Short tandem repeat analysis. Representative D7S280 allele pattern of xenografts and 22R*v1* and CWR-R1 cell lines, along with analysis of six additional loci (fig. S1). An allelic ladder is shown on left and right of gel. (**C**) Quantitative real-time PCR to detect XMRV *env* sequences. Calculated copies per 100 cells are indicated above each bar. (**D**) IAP assay to quantify the amount of mouse DNA present in the xenograft gDNAs.

(the PCR and sequencing coverage are detailed in fig. S3, A and B).

To screen for the presence of endogenous XMRV in mouse strains, we developed an XMRV-specific PCR assay based on sequence differences in the LTR and *gag* leader regions that excluded all known endogenous murine retroviruses (fig. S2). A survey of 45 laboratory mouse strains and 44 wild mice failed to detect XMRV (fig. S4). In a search for proviruses that might contain XMRV-specific sequence features, we found a second previously undescribed endogenous provirus that we named PreXMRV-2 (Fig. 2C). A portion of PreXMRV-2 corresponds to a 1124-nt sequence of an endogenous provirus from the 129X1/SvJ mouse genome (acc. no. AAHY0159188.1) (*6*, *20*). The sequence of PreXMRV-2 (GenBank accession no. FR871850) revealed that *gag*, *pol*, and *env* reading frames are open and can potentially express functional proteins. A 3.6-kb stretch encompassing the *gag* leader region and *gag-pro-pol* differs by one base from the consensus XMRV (99.9% identity); in addition, a ~700-nt region of *env* is 99% identical to XMRV; however, the LTRs and the remaining viral genome differ by 6 to 12% from consensus XMRV. Phylogenetic analysis indicates that PreXMRV-1 is grouped with xenotropic viruses, whereas PreXMRV-2 appears to be a recombinant, grouping with polytropic and modified polytropic viruses for certain stretches of its genome (fig. S5).

We screened 15 mouse strains, which included 12 nude mice, for the presence of XMRV, PreXMRV-1, and PreXMRV-2 using XMRV-specific primers, primers that amplified XMRV or PreXMRV-1, and PreXMRV-2-specific primers (Fig. 2D and fig. S2). None of the mouse strains contained XMRV, and only the Hsd and the NU/NU outbred nude strains contained PreXMRV-1 (Fig. 2D and fig. S6). Six of the 15 mouse strains contained PreXMRV-2, but only the NU/NU and Hsd mice contained both PreXMRV-1 and PreXMRV-2 (Fig. 2D and fig. S6). It should be noted that, because the Hsd and the NU/NU are outbred strains, individual mice differ in their endogenous proviruses. NU/NU mice showed variation in the presence of these two endogenous proviruses, and two out of five animals tested contained both (fig. S6). The 22R*v1* cell line contained only XMRV as confirmed by sequence analysis; however, the CWR-R1 cell line contained both XMRV and PreXMRV-1. The CWR-R1 cell line has been reported to contain contaminating mouse cells (*21*) (and see IAP signal in Fig. 2D), which is likely to be the source of the PreXMRV-1 sequences.

We used the same specific primer sets to determine the distribution of XMRV, PreXMRV-1, and PreXMRV-2 in early and late xenografts (Fig. 2E). None of the early xenografts (736, 777, 9216R, 9218R, 8R, and 8L), but all of the late xenografts (2152, 2524, 2272, and 2274) and both cell lines were positive for XMRV. The primers



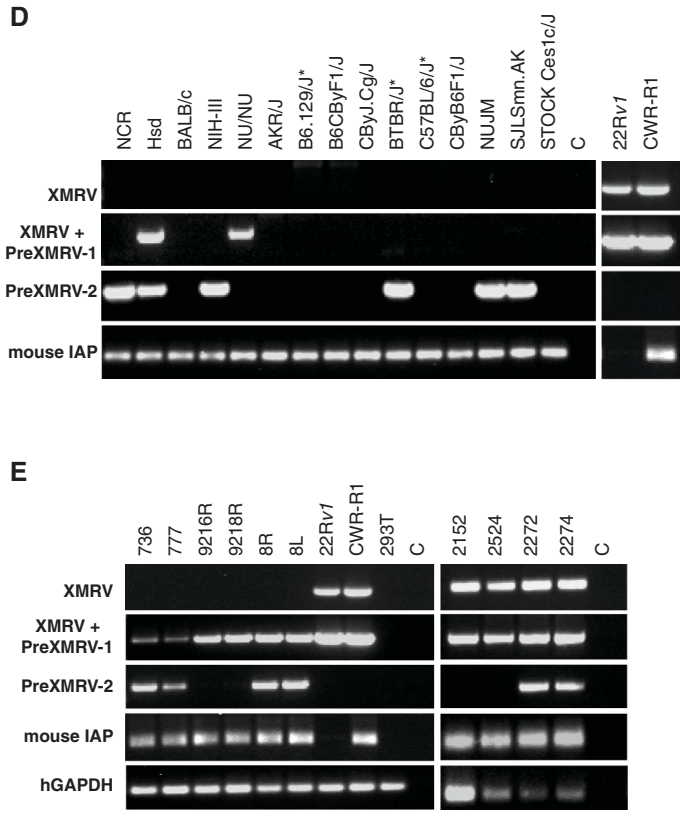


Fig. 2. PCR and sequencing analysis of XMRV and XMRV-related sequences from xenografts, cell lines, and nude mouse strains. Using specific primer sets (fig. S2), cloned PCR products from the xenografts, 22R*v1*, CWR-R1, or mouse strains were sequenced. Approximate length and location of sequences determined from samples that were positive for XMRV (**A**), PreXMRV-1 (**B**), and PreXMRV-2 (**C**) are shown as red bars beneath each provirus. Details of primers and numbers of cloned products sequenced are shown in figs. S2 and S3. Hypermut plots (see fig. S3 for details), which indicate nucleotide mismatches relative to XMRV as color-coded vertical lines, are shown for PreXMRV-1 (B) and PreXMRV-2 (C), together with the percent identity to consensus XMRV for different regions of each provirus [nucleotide numbers refer to the 22R*v1* XMRV sequence (FN692043)]. PreXMRV-1 has a 16-nt deletion (Δ16) in *gag* and a frameshift (fs) in *pol*, which make it replication-defective, whereas PreXMRV-2 *gag*, *pol*, and *env* reading frames are open. (**D**) Mouse strains and (**E**) xenograft and PC cell lines were analyzed by PCR for the presence of XMRV, PreXMRV-1, and PreXMRV-2. Mouse IAP and human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) serve as positive controls for the presence of mouse and human DNA, respectively. For both (D) and (E), the primer set used to detect PreXMRV-1 can also detect XMRV. For ease of comparison, the 22R*v1* and CWR-R1 gel lanes from (E), which were run in parallel, are duplicated in (D). DNAs in (D) and (E) were all amplified with the same PCR primer master mix. †We previously determined the full-length sequence of XMRV from 22R*v1* cells (*8*). Δgap refers to the 24-bp deletion in the *gag* leader characteristic of XMRV. All mouse strains shown in (D) are nudes except for those indicated with an asterisk (*).

used to detect PreXMRV-1 could also detect XMRV; sequencing analysis of the PCR products from all of the early xenografts detected only PreXMRV-1, but both XMRV and PreXMRV-1 were detected from the late xenografts 2524 and 2274 (Fig. 2B). Amplification with PreXMRV-2–specific primers revealed the presence of this provirus in early xenografts 736, 777, 8R, and 8L and late xenografts 2272 and 2274 (Fig. 2, C and E, and fig. S3C). The variable detection level of PreXMRV-2 in the late xenografts could be due to individual differences in the outbred mice, and by extension, in the mouse DNA in these samples.

Comparison of the PreXMRV-1 and PreXMRV-2 sequences revealed that the regions of near identity to XMRV are reciprocal and largely nonoverlapping. We therefore hypothesized that recombination between these two retroviruses resulted in the formation of XMRV. As shown in Fig. 3A, reverse transcriptase template–switching events during minus-strand DNA synthesis can form a recombinant that is essentially identical to the sequences of all of the XMRVs reported to date and that differs from the consensus XMRV by only four nucleotides. The six switching events occurred in 20- to 73-nt stretches that are identical in PreXMRV-1 and PreXMRV-2 (red numbers in Fig. 3A, and fig. S7A). Of the four nucleotide differences between the predicted recombinant and consensus XMRV, only the A>G change at position 790 results in a conservative valine-to-isoleucine amino acid substitution; the other three substitutions are silent. The 22R*v1* and CWR-R1 cell lines, as well as VP42, have an A at position 790, whereas all other XMRV isolates have a G at position 790. The insertion of an A at position 8092 occurred within a run of six adenines; frameshift mutations commonly occur in such homopolymers during retroviral replication (*22*). A comparison of the predicted recombinant to the available XMRV sequences is shown in fig. S7B. The available XMRV sequences all have the same six recombination junctions predicted in the hypothetical recombinant and differ from the consensus XMRV by 3 to 14 nt. These differences may be the result of errors during PCR or sequencing or from mutations during the passage of XMRV in another cell line. Phylogenetic analysis supports the predicted recombinant virus as the precursor of the virus in the CWR22 xenografts, the 22R*v1* and CWR-R1 cell lines, and all XMRVs isolated and sequenced from patients (Fig. 3B) (*23*).

Our findings indicate that virus derived from two previously undescribed murine endogenous retroviruses, PreXMRV-1 and PreXMRV-2, most likely underwent retroviral recombination to generate XMRV during in vivo passaging of the CWR22 xenograft in nude mice. Both parental endogenous proviruses were present in some of the nude mouse strains used for in vivo passaging of the xenografts; therefore, there were opportunities for this recombinant to form and spread in the tumor cells that were the progenitors of the 22R*v1* and CWR-R1 cell lines. Only six template-switching events, which is close to the average of four template switches per replication cycle (*24*), are needed to generate a recombinant that is both essentially identical and ancestral to all XMRV sequences characterized to date from cell lines and patients (Fig. 3B). We have estimated the probability that the exact set of template-switching events occurred independently is 1.3×10^{-12} (fig. S8) (*23*), which makes it very likely that contamination of human samples with XMRV originating from the relapsed CWR22 xenografts or either of the two cell lines, perhaps through other intermediate cell lines, contributed to its reported association with PC and CFS. Our results and conclusions relate to XMRV detection by isolation of virus of this specific sequence (*1*–*3*) and do not directly address detection of antibodies or viral antigens (*25*, *26*) or PCR detection of related but distinct MLV sequences (*27*). We note, however, that most "XMRV-specific" PCR assays may detect PreXMRV-1 or -2 proviruses in contaminating mouse DNA and that specific detection of XMRV requires the use of primers that flank a crossover site.

The alternative possibility is that recombination between PreXMRV-1 and PreXMRV-2 occurred during mouse evolution, giving rise to an endogenous XMRV provirus that is present in mice and can occasionally infect humans. We think this possibility is remote because analysis of the early xenografts, which contained contaminating nude mouse DNAs, failed to detect XMRV. Furthermore, we were unable to detect XMRV in a screen of 89 inbred and wild-derived mouse strains including 17 individual nude mice (fig. S4) (*23*).

We conclude that XMRV was generated as a result of a unique recombination event between two endogenous MLVs that took place around 1993–1996 in a nude mouse carrying the CWR22 PC xenograft. Because the probability that the same recombination event could occur independently by random chance is essentially negligible,

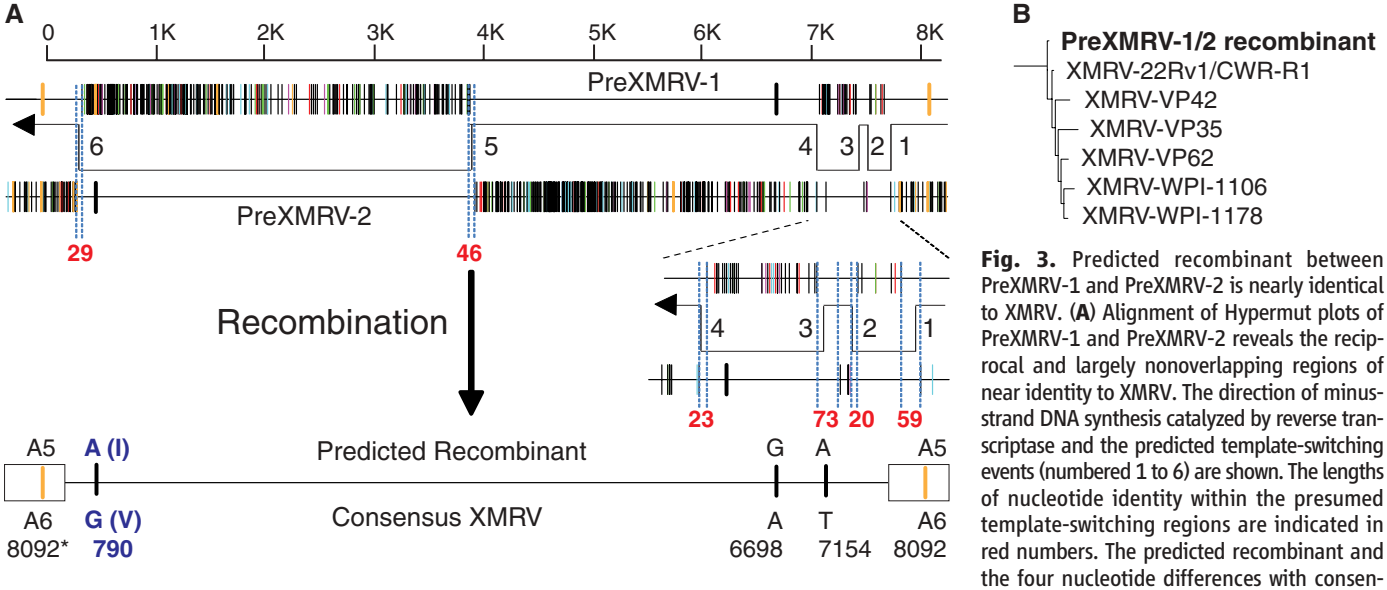


Fig. 3. Predicted recombinant between PreXMRV-1 and PreXMRV-2 is nearly identical to XMRV. (**A**) Alignment of Hypermut plots of PreXMRV-1 and PreXMRV-2 reveals the reciprocal and largely nonoverlapping regions of near identity to XMRV. The direction of minus-strand DNA synthesis catalyzed by reverse transcriptase and the predicted template-switching events (numbered 1 to 6) are shown. The lengths of nucleotide identity within the presumed template-switching regions are indicated in red numbers. The predicted recombinant and the four nucleotide differences with consensus XMRV are shown. The nucleotide numbers refer to numbers of the 22R*v1* XMRV (acc. no. FN692043). Note that nucleotide 8092 is within the U3 region and is present in both LTRs (boxes). A5 and A6 refer to homopolymeric runs of five and six adenines, respectively. The A>G change at 790 results in an isoleucine (I) to valine (V) substitution. (**B**) Phylogenetic tree of all full-length XMRV sequences to date and the predicted recombinant implicates the predicted recombinant as the ancestor of all sequenced XMRV isolates. The tree shown is an enlargement of the XMRV-specific portion of the complete endogenous MLV tree (fig. S5A) (*23*).

any XMRV isolates with the same or nearly the same sequences identified elsewhere originated from this event (*23*).

References and Notes

1. R. Schlaberg, D. J. Choe, K. R. Brown, H. M. Thaker, I. R. Singh, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16351 (2009).
2. A. Urisman *et al.*, *PLoS Pathog.* **2**, e25 (2006).
3. V. C. Lombardi *et al.*, *Science* **326**, 585 (2009).
4. A. C. Van der Kuyl, M. Cornelissen, B. Berkhout, *Front. Microbiol.* **1**, 147 (2011).
5. R. A. Weiss, *BMC Biol.* **8**, 124 (2010).
6. S. Hué *et al.*, *Retrovirology* **7**, 111 (2010).
7. H. C. Groom, M. W. Yap, R. P. Galão, S. J. Neil, K. N. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5166 (2010).
8. T. Paprotka *et al.*, *J. Virol.* **84**, 5719 (2010).
9. K. Stieler, N. Fischer, *PLoS ONE* **5**, e11738 (2010).
10. H. P. Bogerd, F. Zhang, P. D. Bieniasz, B. R. Cullen, *Virology* **410**, 234 (2011).
11. C. Chaipan *et al.*, *J. Virol.* **85**, 4888 (2011).
12. R. M. Sramkoski *et al.*, *In Vitro Cell. Dev. Biol. Anim.* **35**, 403 (1999).
13. E. C. Knouf *et al.*, *J. Virol.* **83**, 7353 (2009).
14. M. Nagabhushan *et al.*, *Cancer Res.* **56**, 3042 (1996).
15. T. G. Pretlow *et al.*, *J. Natl. Cancer Inst.* **85**, 394 (1993).
16. C. W. Gregory, R. T. Johnson Jr., J. L. Mohler, F. S. French, E. M. Wilson, *Cancer Res.* **61**, 2892 (2001).
17. Materials and methods are available as supporting material on *Science* Online.
18. B. Oakes *et al.*, *Retrovirology* **7**, 109 (2010).
19. M. J. Robinson *et al.*, *Retrovirology* **7**, 108 (2010).
20. V. Courgnaud, J.-L. Battini, M. Sitbon, A. L. Mason, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15666 (2010).
21. A. van Bokhoven *et al.*, *Prostate* **57**, 205 (2003).
22. V. K. Pathak, H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6019 (1990).
23. Further discussion in support of the unique origin of XMRV and its role as the source of XMRV in PC and CFS patients is available as supporting material in *Science* Online.
24. J. Zhuang, S. Mukherjee, Y. Ron, J. P. Dougherty, *J. Virol.* **80**, 6706 (2006).
25. R. S. Arnold *et al.*, *Urology* **75**, 755 (2010).
26. X. Qiu *et al.*, *Retrovirology* **7**, 68 (2010).
27. S. C. Lo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15874 (2010).

Acknowledgments: We thank W. Shao for analysis of MLV diversity, and E. Freed and S. Hughes for helpful discussions. This research was supported in part by the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, NIH. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This work was also supported in part by a Bench-to-Bedside Award to V.K.P., research grant R37 CA 089441 to J.M.C., and R01CA150197 to H.J.K. J.M.C. was a Research Professor of the American Cancer Society with support from the F.M. Kirby Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1205292/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References

1 February 2011; accepted 5 May 2011
Published online 31 May 2011;
10.1126/science.1205292

Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice

Jeremiah J. Faith, Nathan P. McNulty, Federico E. Rey, Jeffrey I. Gordon*

The interrelationships between our diets and the structure and operations of our gut microbial communities are poorly understood. A model community of 10 sequenced human gut bacteria was introduced into gnotobiotic mice, and changes in species abundance and microbial gene expression were measured in response to randomized perturbations of four defined ingredients in the host diet. From the responses, we developed a statistical model that predicted over 60% of the variation in species abundance evoked by diet perturbations, and we were able to identify which factors in the diet best explained changes seen for each community member. The approach is generally applicable, as shown by a follow-up study involving diets containing various mixtures of pureed human baby foods.

Owing to its many roles in human health (*1*–*3*), there is great interest in deciphering the principles that govern the operations of an individual's gut microbiota. Current estimates indicate that each of us harbors several hundred bacterial species in our intestine (*4*, *5*), and different diets lead to large and rapid changes in the composition of the microbiota (*6*, *7*). Given the dynamic interrelationship between diet, the configuration of the microbiota, and the partitioning of nutrients in food to the host, inferring the rules that govern the microbiota's responses to dietary ingredients represents a challenge (*8*).

Gnotobiotic mice colonized with simple, defined collections of sequenced representatives of the various phylotypes present in the human gut microbiota provide a simplified in vivo model system in which metabolic niches, host-microbe, and microbe-microbe interactions can be examined by using a variety of techniques (*9*–*12*). These studies have focused on small communities exposed to a few perturbations. We used gnotobiotic mice harboring a 10-member community of sequenced human gut bacteria to model the response of a microbiota to changes in host diet. We aimed to predict the absolute abundance of each species in this microbiota on the basis of knowledge of the composition of the host diet. Furthermore, we wanted to gain insights into the niche preferences of members of the microbiota and to discover how much of the response of the community was a reflection of their phenotypic plasticity.

The 10 bacterial species were introduced into germ-free mice to create a model community with representatives of the four most prominent bacterial phyla in the healthy human gut microbiota (fig. S1A) (*13*). Their genomes encode major metabolic functions that have been identified in anaerobic food webs, including the ability to break down complex dietary polysaccharides not accessible to the host (*Bacteroides thetaiotaomicron*, *Bacteroides ovatus*, and *Bacteroides caccae*); consume oligosaccharides and simple sugars (*Eubacterium rectale*, *Marvinbryantia formatexigens*, *Collinsella aerofaciens*, and *Escherichia coli*); and ferment amino acids (*Clostridium symbiosum* and *E. coli*). We also included two species capable of removing the end products of fermentation: a H_2-consuming, sulfate-reducing bacterium (*Desulfovibrio piger*) and a H_2-consuming acetogen (*Blautia hydrogenotrophica*).

To perturb this community, we used a series of refined diets in which each ingredient represented the sole source of a given macronutrient (casein = protein, corn oil = fat, cornstarch = polysaccharide, and sucrose = simple sugar) and in which the concentrations of these four ingredients were systematically varied (fig. S1, B and C, and table S1). Each individually caged male C57Bl/6J mouse was fed a randomly selected diet, with diet switches occurring every 2 weeks (n = 13 animals; fig. S1D shows the variation of diet presentation between animals). Shotgun sequencing of total fecal DNA allowed us to determine the absolute abundance of each community member, based on assignment of reads to the various species' genomes, in samples obtained from each mouse on days 1, 2, 4, 7, and 14 of a given diet period (*13*).

To predict the abundance of each species in the model human gut microbiome given only knowledge of the concentration of each of the four perturbed diet ingredients, we used a linear model

$$y_i = \beta_0 + \beta_{\text{casein}} X_{\text{casein}} + \beta_{\text{starch}} X_{\text{starch}} + \beta_{\text{sucrose}} X_{\text{sucrose}} + \beta_{\text{oil}} X_{\text{oil}} \quad (1)$$

where y_i is the absolute abundance of species i; X_{casein}, X_{starch}, X_{sucrose}, and X_{oil} are the amounts (in grams per kilogram of mouse diet) of casein,

Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA.

*To whom correspondence should be addressed. E-mail: jgordon@wustl.edu

Recombinant Origin of the Retrovirus XMRV

Tobias Paprotka, Krista A. Delviks-Frankenberry, Oya Cingöz, Anthony Martinez, Hsing-Jien Kung, Clifford G. Tepper, Wei-Shau Hu, Matthew J. Fivash Jr., John M. Coffin and Vinay K. Pathak

Science **333** (6038), 97-101.
DOI: 10.1126/science.1205292originally published online May 31, 2011