
This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of February 3, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/331/6017/593.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2011/01/10/science.1200801.DC1.html>

This article **cites 23 articles**, 7 of which can be accessed free:

<http://www.sciencemag.org/content/331/6017/593.full.html#ref-list-1>

This article appears in the following **subject collections**:

Medicine, Diseases

<http://www.sciencemag.org/cgi/collection/medicine>

Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers

David T. Ting,^{1*} Doron Lipson,^{2*} Suchismita Paul,¹ Brian W. Brannigan,¹ Sara Akhavanfard,¹ Erik J. Coffman,¹ Gianmarco Contino,¹ Vikram Deshpande,¹ A. John Iafrate,¹ Stan Letovsky,² Miguel N. Rivera,¹ Nabeel Bardeesy,¹ Shyamala Maheswaran,¹ Daniel A. Haber^{1†}

Satellite repeats in heterochromatin are transcribed into noncoding RNAs that have been linked to gene silencing and maintenance of chromosomal integrity. Using digital gene expression analysis, we showed that these transcripts are greatly overexpressed in mouse and human epithelial cancers. In 8 of 10 mouse pancreatic ductal adenocarcinomas (PDACs), pericentromeric satellites accounted for a mean 12% (range 1 to 50%) of all cellular transcripts, a mean 40-fold increase over that in normal tissue. In 15 of 15 human PDACs, alpha satellite transcripts were most abundant and HSATII transcripts were highly specific for cancer. Similar patterns were observed in cancers of the lung, kidney, ovary, colon, and prostate. Derepression of satellite transcripts correlated with overexpression of the long interspersed nuclear element 1 (LINE-1) retrotransposon and with aberrant expression of neuroendocrine-associated genes proximal to LINE-1 insertions. The overexpression of satellite transcripts in cancer may reflect global alterations in heterochromatin silencing and could potentially be useful as a biomarker for cancer detection.

Genome-wide sequencing approaches have revealed an increasing set of transcribed noncoding sequences (ncRNA), including “pervasive transcription” by heterochromatic regions of the genome linked to transcriptional silencing and chromosomal integrity (1, 2). In the mouse, heterochromatin is composed of centric (minor) and pericentric (major) satellite repeats that are required for formation of the mitotic spindle complex and faithful chromosome segregation (3), whereas human satellite repeats have been divided into multiple classes with similar functions (4). Accumulation of satellite transcripts in mouse and human cell lines results from DNA demethylation, heat shock, or the induction of apoptosis, and their overexpression has been

associated with genomic instability (5, 6). Stress-induced transcription of satellites in cultured cells has also been linked to the activation of retroelements encoding RNA polymerase activity such as long interspersed nuclear element 1 (LINE-1) (*LITDI*) (7, 8). The global expression of repetitive ncRNAs in primary tumors has not been analyzed owing to the bias of microarray platforms toward annotated coding sequences and the specific exclusion of repeat sequences from standard analytic programs.

We used a next-generation digital gene expression (DGE) method (9) to obtain a comprehensive view of the transcriptome of primary tumors. We first evaluated mouse pancreatic ductal adenocarcinomas (PDACs) generated through pancreas-targeted expression of activated *Kras* and loss of *Tp53* (10). These tumors are histopathological and genetic mimics of human PDAC, which almost universally display mutations in the *KRAS* oncogene and show frequent loss of the *TP53* tumor suppressor gene. Notably, 47% of transcripts sequenced in the first PDAC (468,359 transcripts per million; tpm) were not annotated and mapped to the major mouse satellite, which contributes to

only 0.02 to 0.4% of transcripts in normal pancreas or liver. In the tumor, satellite reads were found in both sense and antisense directions and were absent from purified polyadenylated RNA. The number of transcripts was >100 times that of normal tissue and 3600 times as abundant in the tumor as mRNA transcripts of the *Gapdh* (glyceraldehyde-3-phosphate dehydrogenase) housekeeping gene. We extended DGE analysis to additional mouse tumors with diverse genotypes: Increased satellite expression was noted in 7 of 9 PDACs, 2 of 3 colon cancers, and 2 of 2 lung cancers (range 12,236 to 160,186 tpm) (Fig. 1A and table S1). In primary tumors overexpressing satellites, the composite distribution of all RNA reads among coding, ribosomal, and other nc transcripts differed significantly from that of normal tissues (Fig. 1B), suggesting that the cellular transcriptional machinery is affected by the massive expression of satellites. Genomic amplification of satellites did not account for the exceptional abundance of these transcripts, as determined by next-generation DNA digital copy number variation analysis, implicating transcriptional derepression of heterochromatin as a possible driving mechanism (table S2).

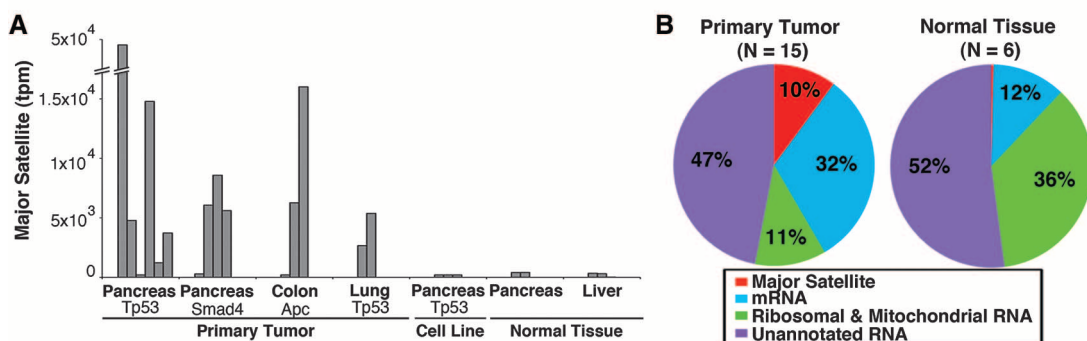
Northern blots of mouse PDACs demonstrated that the major satellite-derived transcripts ranged from 100 base pairs (bp) to 5 kbp (Fig. 2A), consistent with the proposed cleavage of the primary transcript by *Dicer1* (11), whose expression is 2.6 times higher in mouse pancreatic tumors with increased satellite expression ($P = 0.0006$, t test). Immortalized cell lines established from three satellite-overexpressing PDACs displayed minimal expression of satellites (range 173 to 433 tpm), suggesting either negative selection pressure or reestablishment of satellite silencing mechanisms under in vitro conditions. Treatment with 5-azacytidine (AZA) led to massive reexpression of satellites, supporting DNA methylation as a potential mechanism for satellite silencing in vitro (Fig. 2, A and B). Inoculation of an established PDAC cell line (CL3) into nude mice to generate subcutaneous tumors ($n = 5$) resulted in reexpression of satellites, suggesting that these loci become derepressed in vivo (Fig. 2C). Most normal adult mouse tissues, except lung, showed minimal expression of satellites, but the uncleaved 5-kbp satellite tran-

¹Massachusetts General Hospital Cancer Center and Departments of Medicine, Pathology, and Surgery, Harvard Medical School, Boston, MA 02114, USA. ²Helicos BioSciences Corporation, One Kendall Square, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: haber@helix.mgh.harvard.edu

Fig. 1. Massive expression of major satellites in mouse pancreatic tumors. **(A)** Expression of major satellite in primary tumors, cell lines, and normal tissues, presented as transcripts per million of aligned genomic reads. All tumors and cell lines have *Kras*^{G12D}; deleted genes are listed individually (*Tp53*, *Smad4*, and *Apc*). **(B)** Graphical representation of sequence read contributions from major satellites, averaged among all primary tumors versus normal tissues (pancreas and liver). “Unannotated RNA” indicates reads that aligned to the mouse genome, but not to the mouse reference transcriptome.



script was expressed in embryonic tissues (fig. S1). Thus, the aberrant expression of satellites in primary pancreatic tumors does not appear to simply recapitulate an embryonic cell fate, but possibly reflects altered processing of the primary 5-kbp satellite transcript.

RNA in situ hybridization (RNA-ISH) showed high levels of mouse major satellite expression

in all cells within primary tumors and metastases (Fig. 2D). Notably, elevated satellite expression was evident in early preneoplastic low-grade pancreatic intraepithelial neoplasia (PanIN), and it increased further upon transition to high-grade PanIN (Fig. 2E). Clearly defined metastatic lesions to the liver were strongly positive by RNA-ISH, as were small clusters of PDAC cells within the liver parenchyma,

that otherwise would not have been detected by histopathological analysis (fig. S2). Low-level diffuse expression was evident in mouse embryonic liver and lung (fig. S3), but no normal adult or embryonic tissues demonstrated satellite expression comparable to that in tumor cells.

To investigate whether human tumors also overexpress satellite ncRNAs, we extended the DGE analysis to various human malignancies with a particular focus on PDAC. We first measured the total amount of all satellite transcripts: Analysis of 15 PDACs showed a median 21-fold increased expression compared with normal pancreas, but some other normal human tissues also had measurable levels of total satellite expression (fig. S4 and table S3). However, subdivision of human satellites among their multiple classes (4) revealed major differences between tumors and all normal tissues (Fig. 3A). The greatest differential expression in cancer was in the pericentromeric satellite HSATII (mean 2416 tpm; 10.3% of satellite reads), which was undetectable in normal human pancreas and had minimal expression in other normal tissues (131-fold differential expression; Fig. 3, A and B). In contrast, normal tissues had a high representation of GSATII, beta satellite (BSR), and TAR1, although these satellite classes constitute a small minority of satellite reads in pancreatic cancer. The most abundant class of normally expressed human satellites, alpha (ALR) (12), was expressed at 294 tpm in normal human pancreas, but constituted on average 12,535 tpm

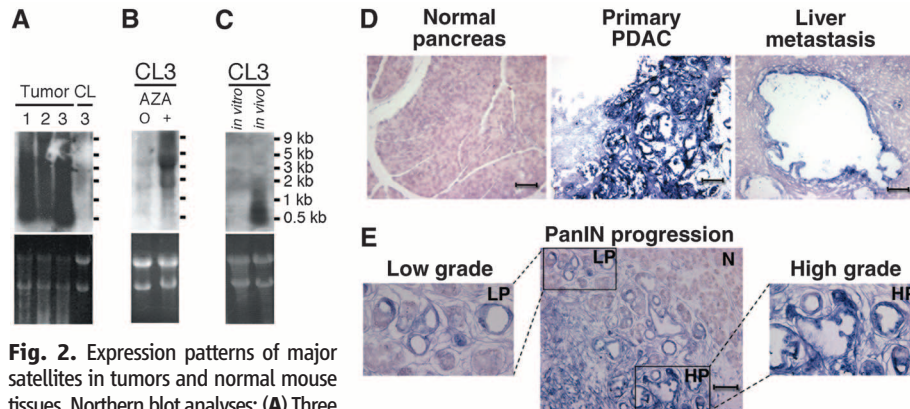
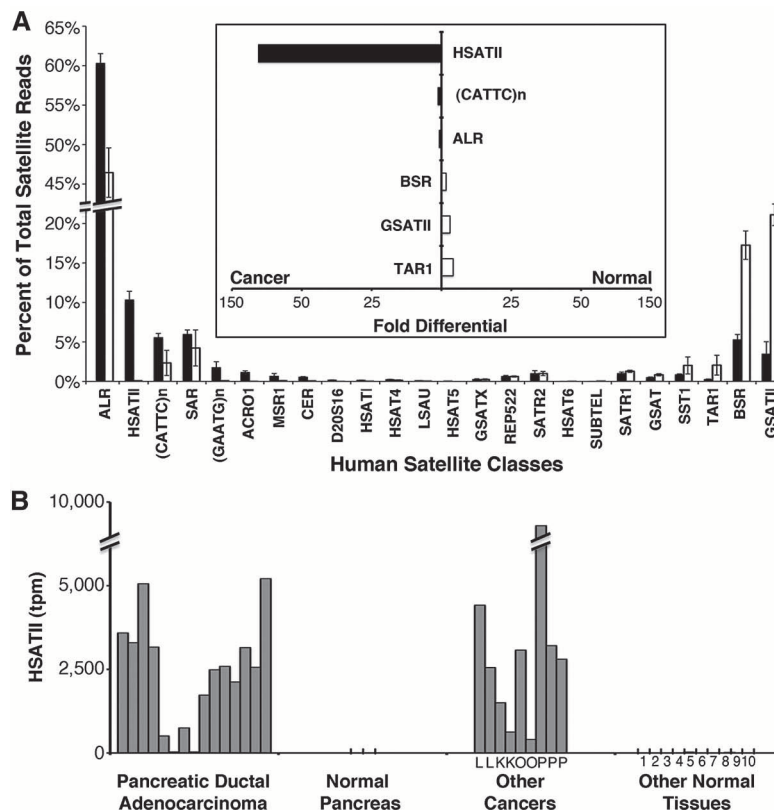
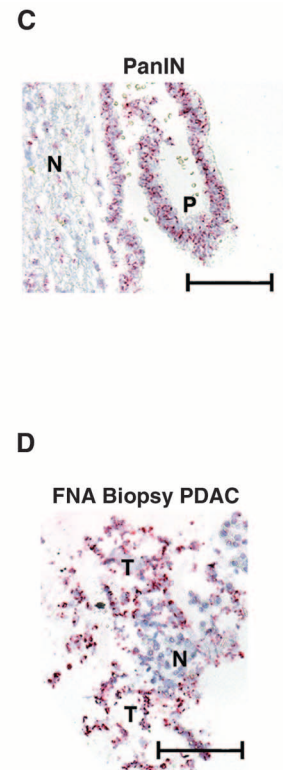


Fig. 2. Expression patterns of major satellites in tumors and normal mouse tissues. Northern blot analyses: (A) Three *Kras*^{G12D}, *Tp53*^{lox/+} pancreatic primary tumors (Tumors 1 to 3) and a stable cell line (CL3) derived from Tumor 3. (B) CL3 before (0) and after (+) treatment with AZA. (C) CL3 cells cultured in vitro and grown as subcutaneous tumors in vivo. (D and E) RNA-ISH with major satellite probe (purple stain): (D) Normal pancreas, primary PDAC, and liver metastasis. (E) Preneoplastic low-grade PanIN (LP) lesion adjacent to high-grade PanIN (HP) and normal pancreas (N). Higher magnification (400×) of inset low-grade (left) and high-grade (right) PanIN lesions. All images are at 200× magnification (scale bar, 100 μm).

Fig. 3. Overexpression of satellites in human cancers. (A) Breakdown of satellite classes as a percentage of total satellites in human PDAC (black, *n* = 15) and normal human tissues (white, *n* = 12). Satellites are ordered from highest in tumors to highest in normal tissue (left to right). Error bars represent SEM. Inset (bar graph, center) shows the differential expression of selected satellite classes enriched in cancer (left, black bars) or normal tissue (right, white bars). (B) HSATII expression in human PDAC, normal pancreas, other cancers (L, lung; K, kidney; O, ovarian; P, prostate), and normal human tissues (1, fetal brain; 2, adult brain; 3, colon; 4, fetal liver; 5, adult liver; 6, lung; 7, kidney; 8, placenta; 9, prostate; and 10, uterus) quantitated by DGE. Satellite expression is shown as transcripts per million (tpm) aligned to human genome. (C and D) RNA-ISH with HSATII probe (red stain): (C) human PanIN (P) and normal adjacent tissue (N). (D) EUS-FNA biopsy of confirmed tumor (T) and normal adjacent tissue (N). All images are at 200× magnification (scale bar, 100 μm).



All images are at 200× magnification (scale bar, 100 μm).



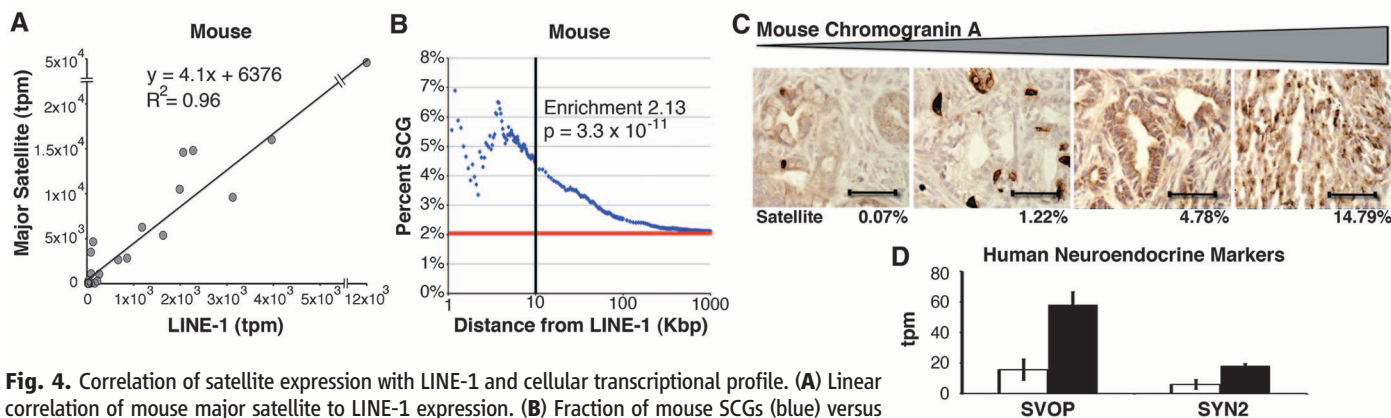


Fig. 4. Correlation of satellite expression with LINE-1 and cellular transcriptional profile. **(A)** Linear correlation of mouse major satellite to LINE-1 expression. **(B)** Fraction of mouse SCGs (blue) versus predicted (red) transcriptional start sites within a given distance of a LINE-1. Enrichment calculations were done at a distance of 10 kbp (black line). **(C)** Immunohistochemistry of mouse PDAC for the neuroendocrine marker chromogranin A. Tumors are depicted as a function of increasing chromogranin A staining (brown), with the relative level of major satellite expression noted for each tumor (bottom; percentage of all transcripts). Images are at 200 \times magnification (scale bar, 100 μ m). **(D)** Differentially expressed ($Q < 0.05$) neuroendocrine genes in human PDACs with high (black) versus low (white) ALR satellite levels. Error bars represent SEM.

in PDACs (60.3% of satellite reads; 43-fold differential expression). Thus, whereas the overexpression of human ALR was comparable to that of mouse major satellites, the less abundant HSATII showed exceptional specificity for human PDAC. High levels of HSATII were also observed in other human cancers, including lung (2 of 2), kidney (2 of 2), ovarian (2 of 2), and prostate (3 of 3), indicating that this may be a shared feature of various carcinomas (mean 2820 tpm; Fig. 3B).

RNA-ISH analysis of human tissues showed differential expression of HSATII in PDAC and PanIN ($n = 4$) compared to normal adjacent tissue as well as in chronic pancreatitis ($n = 8$) (Fig. 3C and fig. S5). When we applied this assay to clinical samples [endoscopic ultrasound-guided fine-needle aspirates (EUS-FNA) of pancreatic masses], HSATII-positive cells were identified in 10 of 10 cases confirmed to have pancreatic cancer at the time of surgical resection, including two cases in which the FNA histopathology was non-diagnostic (Fig. 3D). These initial results suggest that HSATII merits further study as a potential cancer biomarker.

To identify other transcripts co-regulated with satellites in tumors, we performed linear regression analysis in both mouse (major satellite) and human (ALR satellite) (fig. S6). Using a linear correlation cutoff of $R > 0.85$, we created two sets of highly correlative genes (mouse: 297 genes, table S4; human: 539 genes, table S5), which we refer to as satellite correlated genes (SCGs). Mouse and human SCGs were enriched for transposable elements, with the autonomous retrotransposon LINE-1 having the highest expression level in tumors (Fig. 4A). In addition to transposons, a subset of cellular mRNAs showed high correlation with the expression of satellites across diverse tissues. Absence of a shared transcriptional silencing mechanism may contribute to derepression of both LINE-1 and satellites, but the increased expression of diverse mRNAs is less readily explained. LINE-1 insertion upstream of transcriptional start sites of cellular transcripts has recently been implicated in gene regulation (13–15), leading

us to test the proximity of genomic LINE-1 insertions to the SCGs. In mouse, there was a marked correlation between SCGs and their distance to LINE-1 genomic insertions (Fig. 4B). A similar measurable effect was evident with human SCGs, albeit dampened most likely by the heterogeneity of LINE-1 insertions in the human genome (16–18) (fig. S7). Together, these observations suggest that tumor-associated derepression of satellites is highly correlated with increased expression of LINE-1, along with a subset of cellular genes in close proximity to this retrotransposon.

Of cellular transcripts that constitute SCGs, 190 of 297 mouse SCGs and 206 of 539 human SCGs are recognized by the DAVID gene ontology program (19, 20); in both species, the transcripts are highly enriched for genes implicated in neural cell fates and germ or stem cell pathways (table S6). Neuroendocrine differentiation has been described in a variety of epithelial malignancies, including pancreatic cancer (21), and it is correlated with increased aggressiveness in prostate cancer (22). In mouse PDACs, we observed a marked correlation between the level of satellite expression and the number of carcinoma cells staining for the neuroendocrine marker chromogranin A (Fig. 4C), whereas in human PDACs, the neuroendocrine markers synaptic vesicle 2-related protein (SVOP) and synapsin 2 (SYN) were associated with high ALR satellite expression (Fig. 4D and table S7). Together these data suggest that a global alteration in expression of heterochromatic ncRNAs may affect a known cellular differentiation program implicated in cancer.

In summary, we have identified the massive generation of bidirectional ncRNAs from the major satellite in mouse tumor models and from ALR and HSATII satellites in human pancreatic and other epithelial cancers. The discovery of satellite repeat overexpression was made possible by the development of next-generation DGE approaches, which provide a quantitative and sequence-specific measure of highly repetitive sequences that are excluded from traditional analytic programs. Indeed, BLAST sequence matching of satellite sequences

in both mouse and human tumors first identified sequences in the recently completed parasite genomes (see supporting online text). Although further analyses are required to explore the mechanism and consequences of aberrant expression of satellites in cancer tissues, we hypothesize that it likely results from a general derepression of chromosomal marks affecting both satellites and LINE-1 retrotransposons, with proximity to LINE-1 activation affecting the expression of cellular genes enriched for neuroendocrine specification. Current evidence indicates that both DNA methylation and histone H3 lysine 9 (H3K9) trimethylation are critical for the maintenance of satellite repression (14) and that dysregulation of these epigenetic marks is linked to carcinogenesis. Targeted DGE analysis of all known epigenetic regulators (23) in mouse and human tumors showed distinct expression patterns, but no single consistent abnormality (tables S8 to S10). Finally, the potential importance of satellite and LINE-1 deregulation as a consistent biomarker in diverse epithelial cancers merits further clinical testing.

References and Notes

1. J. Berretta, A. Morillon, *EMBO Rep.* **10**, 973 (2009).
2. A. Jacquier, *Nat. Rev. Genet.* **10**, 833 (2009).
3. M. Guenatri, D. Bailly, C. Maisson, G. Almouzni, *J. Cell Biol.* **166**, 493 (2004).
4. J. Jurka *et al.*, *Cytogenet. Genome Res.* **110**, 462 (2005).
5. H. Bouzina-Segard, A. Guais, C. Francastel, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8709 (2006).
6. R. Valgardsdottir *et al.*, *Nucleic Acids Res.* **36**, 423 (2008).
7. D. Ugarkovic, *EMBO Rep.* **6**, 1035 (2005).
8. D. M. Carone *et al.*, *Chromosoma* **118**, 113 (2009).
9. D. Lipson *et al.*, *Nat. Biotechnol.* **27**, 652 (2009).
10. N. Bardeesy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5947 (2006).
11. C. Kanellopoulou *et al.*, *Genes Dev.* **19**, 489 (2005).
12. T. Okada *et al.*, *Cell* **131**, 1287 (2007).
13. T. Kuwabara *et al.*, *Nat. Neurosci.* **12**, 1097 (2009).
14. J. A. Bailey, L. Carrel, A. Chakravarti, E. E. Eichler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6634 (2000).
15. D. E. Montoya-Durango *et al.*, *Mutat. Res.* **665**, 20 (2009).
16. C. R. Beck *et al.*, *Cell* **141**, 1159 (2010).
17. C. R. Huang *et al.*, *Cell* **141**, 1171 (2010).
18. R. C. Iskow *et al.*, *Cell* **141**, 1253 (2010).

19. G. Dennis Jr. *et al.*, *Genome Biol.* **4**, P3 (2003).
20. W. Huang da, B. T. Sherman, R. A. Lempicki, *Nat. Protoc.* **4**, 44 (2009).
21. E. Tezel, T. Nagasaka, S. Nomoto, H. Sugimoto, A. Nakao, *Cancer* **89**, 2230 (2000).
22. L. Cindolo, M. Cantile, F. Vacherot, S. Terry, A. de la Taille, *Urol. Int.* **79**, 287 (2007).
23. P. A. Cloos, J. Christensen, K. Agger, K. Helin, *Genes Dev.* **22**, 1115 (2008).
24. This work was supported by a Pancreatic Cancer Action Network–American Association for Cancer Research

Fellowship and the Warsaw Institute for Pancreatic Cancer Research (D.T.T.); Fond.Veronesi (G.C.); Howard Hughes Medical Institute (D.A.H. and M.N.R.); and National Cancer Institute CA129933 (D.A.H.). We thank T. Raz, P. Kapranov, E. Giladi, and J. Whetstine for helpful discussions and K. Haigis and K. Wong for providing mouse colon and lung tumors, respectively. Massachusetts General Hospital and the authors (D.A.H., D.L., S.M., D.T.T.) have filed a patent application relating to detection of satellite and LINE sequences in human cancers.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1200801/DC1
 Materials and Methods
 SOM Text
 Figs. S1 to S7
 Tables S1 to S10
 References

28 September 2010; accepted 4 January 2011
 Published online 13 January 2011;
 10.1126/science.1200801

A Cryptic Subgroup of *Anopheles gambiae* Is Highly Susceptible to Human Malaria Parasites

Michelle M. Riehle,^{1,2*} Wamdaogo M. Guelbeogo,^{3*} Awa Gneme,³ Karin Eiglmeier,¹ Inge Holm,¹ Emmanuel Bischoff,¹ Thierry Garnier,¹ Gregory M. Snyder,² Xuanzhong Li,⁴ Kyriacos Markianos,⁴ N’Fale Sagnon,³ Kenneth D. Vernick^{1,2,†}

Population subgroups of the African malaria vector *Anopheles gambiae* have not been comprehensively characterized owing to the lack of unbiased sampling methods. In the arid savanna zone of West Africa, where potential oviposition sites are scarce, widespread collection from larval pools in the peridomestic human habitat yielded a comprehensive genetic survey of local *A. gambiae* population subgroups, independent of adult resting behavior and ecological preference. A previously unknown subgroup of exophilic *A. gambiae* is sympatric with the known endophilic *A. gambiae* in this region. The exophilic subgroup is abundant, lacks differentiation into M and S molecular forms, and is highly susceptible to infection with wild *Plasmodium falciparum*. These findings might have implications for the epidemiology of malaria transmission and control.

Much of the current genetic knowledge of African malaria vector populations is based on collection and analysis of indoor house-resting (endophilic) mosquitoes. Indoor capture of mosquitoes is often regarded as an unbiased collection method for epidemiologically important vectors of human malaria, including *A. gambiae* sensu stricto (ss), because they are thought to be “naturally endophilic” (1). Outdoor-resting (exophilic) mosquitoes can contribute to malaria transmission but are underrepresented or absent from indoor collections, particularly if they also bite outdoors (2–7). Collection methods for exophilic mosquitoes are much less efficient than for indoor-resting mosquitoes, involving variants of (i) artificial resting sites such as pits, boxes, or pots; (ii) manual aspiration from vegetation and holes; or (iii) capture of mosquitoes landing on animal or human bait, now largely proscribed be-

cause of the risk of infection of human collectors (2). Further highlighting the challenge in sampling exophilic mosquitoes, resting sites thought to harbor large *A. gambiae* populations during the dry season have resisted detection for decades (8–10).

Little is known about the population genetics of resting site choice, but the Garki malaria control project in Nigeria in the 1970s failed largely because preexisting genetically controlled exophilic behavior allowed persistent malaria transmission by outdoor-resting mosquitoes, despite widespread indoor residual insecticide spraying (3, 11).

Genetic division of *A. gambiae* populations into subgroups allows fine ecological partitioning by the species, mediating the expansion of malaria transmission spatially and temporally. Chromosome inversion polymorphisms define certain subgroups (10, 12). Frequency of the 2La inversion follows a geographic cline from the humid Central African forest, where the wild-type 2La+ allele is fixed, north to the arid West African savanna, where the inverted 2La allele is fixed (12). One likely phenotype of the 2La inversion is adaptation to aridity. Another example of niche expansion of *A. gambiae* by genetic subdivision is represented by two genetically diverged molecular forms, termed M and S (13). These molecular forms are detected by assays for fixed nucleotide differences on the X chromosome (Molecular Form Diagnostic SNPs, MFDS), and display other fixed single-nucleotide polymorphisms (SNPs) in genomic “speciation islands”

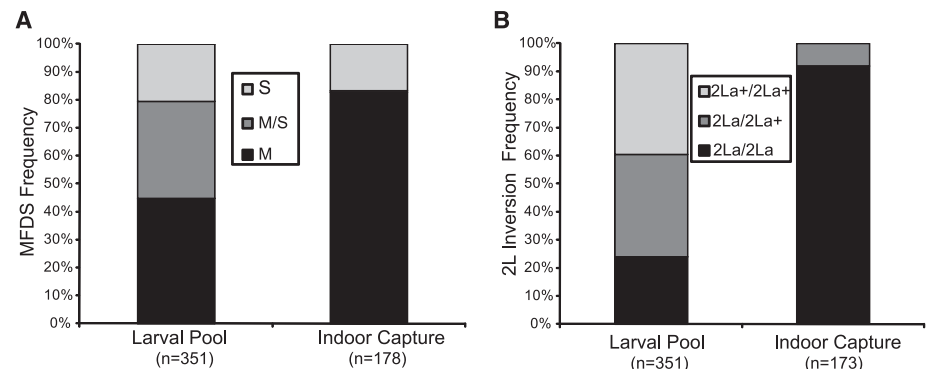


Fig. 1. Comprehensive sampling of *A. gambiae* populations in the West African Sudan-Savanna zone reveals unexpected genotypes. Histograms compare populations of indoor-resting and larval pool collections of *A. gambiae* ss. Larval pools are a limiting resource in the arid zone, and larval pool sampling represents an unbiased survey of all population subgroups, whereas indoor house collections sample behaviorally endophilic mosquitoes. Frequencies of (A) molecular form M and S diagnostic SNP alleles (MFDS) and (B) 2La inversion alleles are significantly different in larval pool and indoor-resting mosquito collections (for MFDS, $\chi^2 = 89.669$, $P = 3E-20$; for 2La inversion, $\chi^2 = 178.533$, $P = 1.7E-39$). The wild-type 2La+ chromosome allele and hybrids of the MFDS occur at unexpectedly high frequency in larval collections as compared with indoor-resting mosquito collections, which suggests population genetic substructure and behavioral partitioning. All samples are *A. gambiae* ss by molecular and cytogenetic species assays, collected ≤ 1 km from human houses at three independent geographic sites across a 400-km transect in Burkina Faso (map, fig. S1) during the 2007 and 2008 malaria transmission seasons.

¹Institut Pasteur, Unit of Insect Vector Genetics and Genomics, Department of Parasitology and Mycology, CNRS Unit of Hosts, Vectors and Pathogens (URA3012), Department of Environment and Sustainable Development (EDD), 28 rue du Docteur Roux, Paris 75015, France. ²University of Minnesota, Department of Microbiology, 1500 Gortner Avenue, Saint Paul, MN 55108, USA. ³Centre National de Recherche et de Formation sur le Paludisme, 1487 Avenue de l’Oubritenga, 01 BP 2208 Ouagadougou, Burkina Faso. ⁴Program in Genomics, Children’s Hospital Boston, Harvard Medical School, 3 Blackfan Street, Boston, MA 02115, USA.

*These authors contributed equally to this work.
 †To whom correspondence should be addressed. E-mail: kenneth.vernick@pasteur.fr