

## TUMOR EVOLUTION

# Origins of lymphatic and distant metastases in human colorectal cancer

Kamila Naxerova,<sup>1,2\*</sup> Johannes G. Reiter,<sup>3</sup> Elena Brachtel,<sup>4</sup> Jochen K. Lennerz,<sup>4</sup> Marc van de Wetering,<sup>5,6</sup> Andrew Rowan,<sup>7</sup> Tianxi Cai,<sup>8</sup> Hans Clevers,<sup>5,6</sup> Charles Swanton,<sup>7,9</sup> Martin A. Nowak,<sup>3,10</sup> Stephen J. Elledge,<sup>2,11</sup> Rakesh K. Jain<sup>1</sup>

The spread of cancer cells from primary tumors to regional lymph nodes is often associated with reduced survival. One prevailing model to explain this association posits that fatal, distant metastases are seeded by lymph node metastases. This view provides a mechanistic basis for the TNM staging system and is the rationale for surgical resection of tumor-draining lymph nodes. Here we examine the evolutionary relationship between primary tumor, lymph node, and distant metastases in human colorectal cancer. Studying 213 archival biopsy samples from 17 patients, we used somatic variants in hypermutable DNA regions to reconstruct high-confidence phylogenetic trees. We found that in 65% of cases, lymphatic and distant metastases arose from independent subclones in the primary tumor, whereas in 35% of cases they shared common subclonal origin. Therefore, two different lineage relationships between lymphatic and distant metastases exist in colorectal cancer.

The spread of cancer cells from the primary tumor to regional lymph nodes is one of the most important factors predicting survival in patients with epithelial cancers (1). Lymph node metastasis uniformly associates with worse outcomes in breast (2), prostate (3), lung (4), and colorectal cancer (5), the most frequent cancers in the U.S. population. In colorectal carcinoma, the presence of cancer cells in tumor-draining lymph nodes defines stage III disease and triggers administration of adjuvant chemotherapy (6). The 5-year survival for patients with stage II (no lymph node metastases) is 82.5%, in contrast to 59.5% for patients with stage III disease (7).

In most patients, lymph node metastasis is not the cause of death but is correlated with spread to vital organs (8). The association between lymphatic and distant metastasis has been known for at least 150 years (9) and, together with the observation that lymph node disease often precedes systemic disease, has engendered the view

that affected lymph nodes may give rise to distant metastases (10–12). The concept of such a sequential progression or metastatic cascade (13), in which the primary tumor (T) seeds lymph node metastases (N) that in turn seed distant metastases (M), provides a mechanistic basis for the TNM staging system. A corollary of the sequential-progression model is that surgical resection of positive lymph nodes will reduce recurrence rates. Indeed, resection of regional lymph nodes has been performed for more than 100 years (14). More recently, a number of clinical trials have shown that lymph node removal does not always improve patient survival (15). These findings have inspired the alternative view that lymph node metastases do not give rise to distant metastases (16) and suggest that treatment strategies may need to be reevaluated (17).

Given its potential impact on patient care, a better understanding of the evolutionary relationship between lymph node and distant metastases is critical. We still do not know whether a single metastatic subclone evolves in the primary tumor, subsequently spreading to lymph nodes and distant sites (18–21), or whether multiple subclones in the primary tumor independently seed lymphatic and distant metastases (22–24). Here we begin to examine these questions by studying the evolutionary history of colorectal cancer metastases.

## Insertion and deletion mutations (indels) in hypermutable DNA enable reconstruction of tumor phylogenies

We conducted a systematic review of 1373 patient records and diagnostic materials at Massachusetts General Hospital (MGH) (fig. S1) and initially identified 19 colorectal cancer patients for whom formalin-fixed and paraffin-embedded samples from primary tumors, lymph nodes,

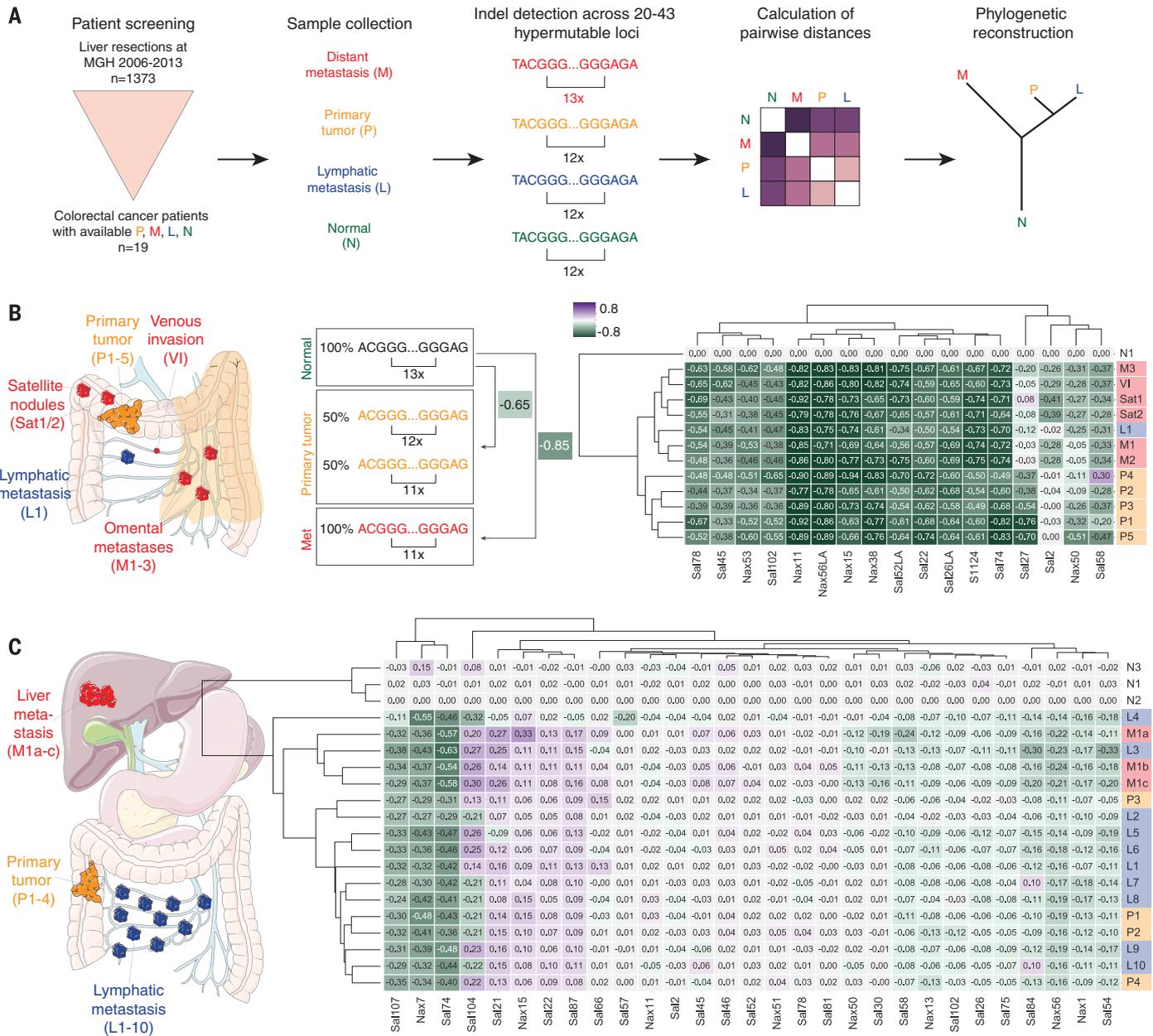
and distant metastases were available (Fig. 1A). We collected multiple tumor regions for each patient (mean 12.6, range 7 to 29) for a total of 239 samples [92 primary tumor biopsies, 59 lymph node metastases, 52 distant metastases, 36 normal tissue (germline) samples] (table S1). Of the 19 patients, 17 had liver metastases, 1 had an ovary metastasis, and 1 had multiple metastases in the omentum.

To trace the evolution of these cancers, we used a methodology (25) that leverages indel mutations in hypermutable, noncoding polyguanine repeats (fig. S2). The mutation rate of polyguanine repeats is several orders of magnitude greater than the mutation rate of nonrepetitive DNA (26), making these sequences a rich reservoir of neutral somatic variation. Previous work has shown that indels in polyguanine repeats (27), as well as other microsatellites (28), accurately reconstruct evolutionary events modeled in cell culture. Furthermore, in silico models of polyguanine-tract evolution have demonstrated that revertant or parallel mutations do not notably affect phylogenetic reconstruction accuracy when the number of interrogated markers is larger than 10 (29). These properties make polyguanine tracts attractive tools for phylogenetic analyses. Here, the mutation information from 20 to 43 polyguanine markers was generally sufficient to resolve lineages at our chosen confidence threshold (clade confidence > 70%, figs. S3 and S4). Polyguanine markers were distributed across many chromosomes (table S2). Therefore, any individual chromosomal alteration (gain or loss) would not be expected to substantially influence phylogenetic reconstruction. In total, our data set consisted of 19,541 individual genotypes. We developed a fully automated pipeline for data filtering, noise reduction, and phylogenetic reconstruction (supplementary methods). First, to avoid artifacts created by contamination with normal cells, we implemented rigorous purity criteria, eliminating 11% of specimens from our study (fig. S5 and table S1). For all specimens belonging to the same patient, we then calculated a pairwise distance measure, the Jensen-Shannon distance (JSD) (30), over all polyguanine repeats. This distance reflected how much the samples had genetically diverged and formed the basis of our phylogenetic reconstruction with the neighbor-joining method (31).

Indel mutation patterns in the colorectal cancer cohort differed among patients. Of all observed alterations, 81% were deletions and 19% were insertions. A 4:1 ratio of deletions to insertions has previously been described by us and others (25, 32), indicating that it is an inherent property of polyguanine repeats. However, among individual cancers, we observed a relatively wide range of deletion frequencies, ranging from 45% in patient C69 to 100% in patient C77 (fig. S6). This suggests that additional determinants, such as alterations in specific DNA repair proteins, may contribute to a skewing of mutation patterns in individual tumors. For example, a cancer showing microsatellite instability due to loss of *MLH1* protein expression almost exclusively harbored

<sup>1</sup>Edwin L. Steele Laboratories for Tumor Biology, Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. <sup>2</sup>Division of Genetics, Department of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. <sup>3</sup>Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, USA. <sup>4</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. <sup>5</sup>Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center (UMC) Utrecht, 3584CT Utrecht, Netherlands. <sup>6</sup>Cancer Genomics Netherlands, UMC Utrecht, 3584CG Utrecht, Netherlands. <sup>7</sup>The Francis Crick Institute, London NW1 1AT, UK. <sup>8</sup>Department of Biostatistics, Harvard University, Boston, MA 02115, USA. <sup>9</sup>University College London Cancer Institute, London WC1E 6DD, UK. <sup>10</sup>Department of Mathematics and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. <sup>11</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

\*Corresponding author. Email: naxerova.kamila@mgh.harvard.edu



**Fig. 1. Tracing tumor evolution through indels in hypermutable DNA.** (A) Study design schematic. DNA samples from primary tumor (P), distant metastases (M), lymph node metastases (L), and normal tissue (germline) (N) from 19 colorectal cancer patients were genotyped across 20 to 43 hypermutable polyguanine repeats. The genetic divergence between two samples is the average distance across all markers. The number of consecutive guanines at a hypothetical locus is indicated in the middle panel. Pairwise distances among all samples from a patient were used as input for phylogenetic reconstruction with the neighbor-joining algorithm. (B) Anatomical sketch and raw data example for a

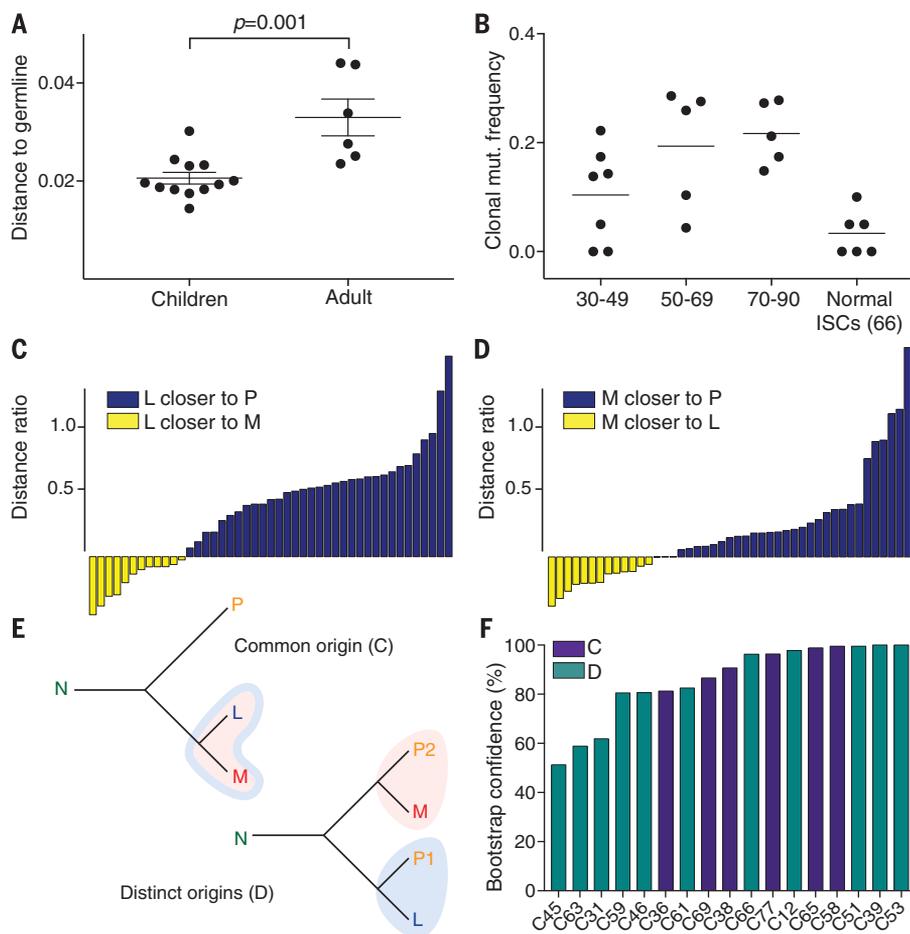
cancer (C38) with microsatellite instability. Mutations can be present in varying percentages of cells within a sample. Therefore, the distance of a tumor sample to the normal reference is a continuous value. The heatmap shows tumor-normal Jensen-Shannon distances across all samples and polyguanine markers. Green, deletions (here, a negative sign indicates deletions); purple, insertions. Note that the heatmap only shows a small part of the full data set for a patient. Pairwise distances between all samples are used for phylogenetic reconstruction. (C) Anatomical sketch and raw data example for a microsatellite-stable cancer (C58). Heatmaps for all patients are provided in fig. S13.

deletions (Fig. 1B) that also were of a considerably larger size than deletions in microsatellite-stable tumors (Fig. 1C and fig. S6). Next, we explored whether accumulation of polyguanine indels is a cancer-related process or whether these mutations can be found in

age-matched normal intestinal stem cells (ISCs). We analyzed DNA from 18 clonal expansions of human ISCs. Stem cell donors for 12 of these expansions were children (ages 4 to 14), and 6 expansions were from a 66-year-old adult (fig. S7). Adult ISCs had diverged significantly

farther from a polyclonal germline reference than ISCs from children (Fig. 2A), suggesting that polyguanine indels accumulate in normal ISCs. We also observed a significant correlation between clonal mutation frequency and patient age in our colorectal cancer cohort (fig. S8).

Downloaded from <http://science.sciencemag.org/> on July 9, 2017



**Fig. 2. Common versus distinct origins of lymph node and distant metastases.** (A) Clonal expansions of single ISCs from children (age < 15 years old,  $n = 12$ ) have fewer polyguanine indels than clonal expansions from a 66-year-old adult ( $n = 6$ ). Data are mean  $\pm$  SEM, two-tailed Student's  $t$  test. (B) Clonal mutation frequency in cancers (defined as  $JSD \geq 0.11$  in 95% of tumor biopsies) is correlated with patient age at diagnosis. Normal ISCs, on average, have fewer polyguanine indels than age-matched cancers, but the two distributions overlap. Lines indicate the mean. (C) Most lymph node metastases are more closely related to the primary tumor than to distant metastases. The plot shows  $d_{(L \text{ to } M)} / d_{(L \text{ to } P)} - 1$ , the distance of each lymph node metastasis (L) to its closest distant metastasis (M), divided by its distance to its closest primary tumor sample (P), minus one. Yellow, closest neighbor is a metastasis; dark blue, closest neighbor is a primary tumor sample. (D) Analogous plot for distant metastases, showing  $d_{(M \text{ to } L)} / d_{(M \text{ to } P)} - 1$ . (E) Classification of patients into cases with common or distinct origins of lymphatic and distant metastases. (F) Bootstrap values reflecting origin-classification confidence for each patient (bootstrap  $n = 1000$ ).

The mean clonal mutation frequency in normal ISCs from the 66-year-old donor was lower than the mean frequency in cancers from age-matched (50- to 69-year-old) patients (Fig. 2B), but with considerable overlap of the two distributions. The baseline polyguanine mutation burden of a colorectal cancer, therefore, partially consists of alterations that are present in all intestinal cells.

### Two distinct patterns of metastatic dissemination exist in colorectal cancer

The main goal of our study was to illuminate the evolutionary relationship between lymphatic and distant metastases. We aimed to sample lymph

nodes as comprehensively as possible and included 91.3% of resected positive nodes in our analysis (Fig. S9, see supplementary methods for a detailed description of lymph node inclusion criteria). We first investigated the genetic distances among lymph node metastases, primary tumor biopsies, and distant metastases. For 33 of 45 (73%) lymph node metastases, the distance to the primary tumor (Fig. 2C) was shorter than the distance to distant metastases, and 31 of 45 (69%) distant metastases had a shorter genetic distance to the primary tumor than to any lymph node metastasis (Fig. 2D). This indicates that both types of metastatic lesions likely originated from distinct subclones in the primary tumor in most

cases. To test this hypothesis, we examined all phylogenetic trees according to formal criteria. Patients were classified into two categories on the basis of tree topology (Fig. 2E). We reasoned that lymph node and distant metastases had a common origin if a patient's tree contained a clade that included at least one lymph node and at least one distant metastasis but no primary tumor samples. Existence of such a branch indicates that both types of metastases were seeded from the same subclone, or that lymph node metastases gave rise to distant metastases. Formally, the reverse—seeding of lymphatic metastases from distant metastases—is also possible. A patient was classified as having distinct origins of lymphatic and distant metastases if no such clade existed. In all distinct origin cases, lymphatic and distant metastases were each more closely related to a primary tumor region than to each other.

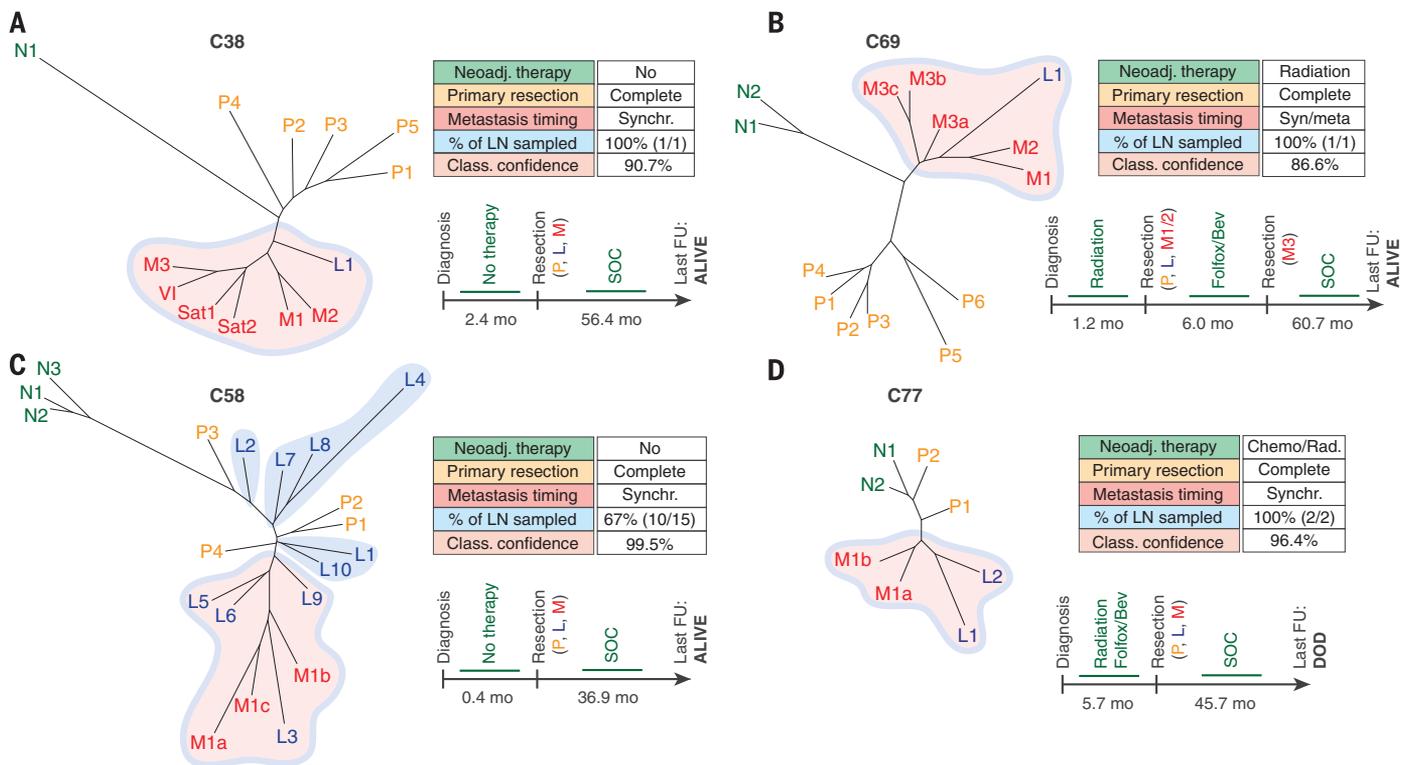
To assess the robustness of each tumor's origin classification, we employed a bootstrapping strategy. We performed repeated random sampling ( $n = 1000$ ) of a tumor's mutation data to determine whether our origin classification was sensitive to changes in a limited number of polyguanine markers (supplementary methods). Of the 17 tumors sampled, 14 (82%) were classified with a bootstrap value above 80% (Fig. 2F), confirming the robustness of our phylogenetic data and classification scheme. We also evaluated our classification by utilizing the unweighted pair-group method with arithmetic mean (UPGMA) (33). (The nature of polyguanine genotyping data suggests the use of distance-based phylogenetic methods; see supplementary methods). Origin classification outcomes did not change for any of our 17 patients, further demonstrating the reliability of our results (all phylogenetic trees are available at Dryad: <http://dx.doi.org/10.5061/dryad.vv53d>).

### Common origin of lymphatic and distant metastases

In 6 out of 17 tumors (35%), we found a common origin of lymphatic and distant metastases. Selected phylogenetic trees with a classification confidence score above 80% for common origin are displayed in Fig. 3, along with pertinent clinical information.

Patient C38's cancer (Fig. 3A and anatomical sketch in Fig. 1B) spread to the omentum and to the mesenteric lymph nodes. Furthermore, several satellite nodules had formed within the colonic epithelium, spatially separated from the primary tumor. We also investigated a piece of tumor that had invaded a vein. Notably, phylogenetic reconstruction showed that all lesions whose formation had depended on cell migration (that is, the satellite nodules, the distant and lymph node metastases, and the tumor within the vein) shared common ancestry, whereas the primary tumor had a divergent genetic profile.

Patient C69's cancer (Fig. 3B) showed a similar pattern, with one metastatic subclone giving rise to several liver metastases and a lymph node metastasis. Polyguanine indels clearly attributed all metastases to the same evolutionary branch,



**Fig. 3. Phylogenetic trees of cancers with a common origin of lymphatic and distant metastases. (A to D)** All trees except C38 [microsatellite-instability (MSI) case] (A) are drawn to scale and were constructed with the neighbor-joining method. Seeding events [internal node (common ancestor) and branches] that gave rise to distant metastases are shaded in red; events that gave rise to lymph node metastases are shaded in blue. Clinical information boxes show whether a patient received neoadjuvant therapy, whether primary tumor resection was complete (all margins

unaffected), whether distant metastases occurred synchronously or metachronously, what percentage of suitable lymph nodes (i.e., those that were large and pure enough, see supplementary methods for details) was sampled, and the origin classification bootstrap value. Timelines summarize treatment and known life span for each patient. Lowercase letters (a, b, c) after sample numbers indicate multiple biopsies from the same tumor mass. VI, venous invasion; Sat, satellite nodule; SOC, standard of care; FU, follow up; DOD, dead of disease.

even though M1, M2, and L1 were resected several months earlier than M3 and the patient received chemotherapy and bevacizumab between surgeries (M and L indicate distant metastases and lymph node metastases, respectively).

Patient C58 (Fig. 3C and the anatomical sketch in Fig. 1C) had widespread metastases to the mesenteric lymph nodes and the liver. Most lymph node metastases were closely related to the primary tumor. However, a distinct subclone had formed in several lymph nodes that were located in close anatomical proximity (L3, L5, L6). The liver metastasis derived from the same subclone found in this lymph node group.

Further examples of patients with a common origin of lymphatic and distant metastases are shown in Fig. 3D and fig. S3. In all common origin cases, tree topologies are consistent with one of two modes of dissemination: The primary tumor seeded lymph node metastases, which in turn seeded distant metastases, with the formal possibility of the reverse; or one genetically distinct ancestor evolved within the primary tumor and subsequently colonized lymph nodes and distant sites. In both scenarios, lymphatic and distant metastases share a common origin.

The common origin category is compatible with the idea of sequential progression and can

explain important clinical observations, such as the well-established correlation between lymphatic and distant disease. In a majority of patients, however, tree topologies indicated independent seeding of lymphatic and distant metastasis from the primary tumor.

**Distinct origins of lymphatic and distant metastases**

Cancers in the distinct origins group, which encompassed 11 out of 17 patients (65%), contained multiple, genetically distinct metastasis ancestors. Figure 4 shows selected phylogenetic trees with a distinct origin–classification confidence score above 80% (the complete set, along with confidence values for each clade, is provided in fig. S4).

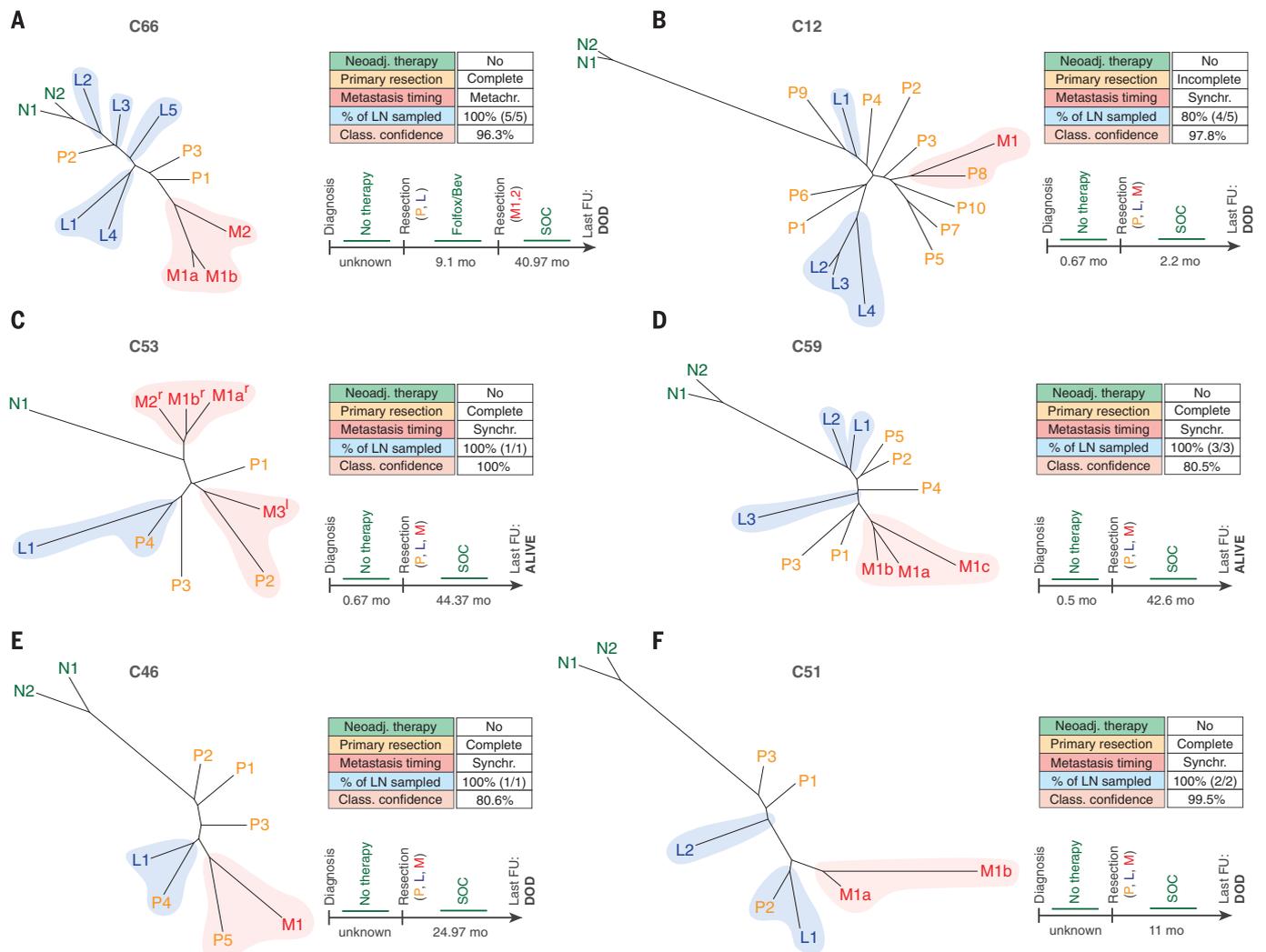
Patient C66’s tumor (Fig. 4A) is a representative example of the distinct origins category. The cancer harbored multiple subclones at different stages of evolution that had seeded genetically distinct metastases. Area P2, for example, was most closely related to lymph node metastasis L3, whereas area P1 was the origin of liver metastases M1 and M2 (P indicates primary tumor). The tree shows that lymph node metastases were seeded continuously throughout the development of the tumor but did not metastasize further. Conversely, the liver metastases arose in

later evolution stages from the genetically most advanced clone. They constitute the terminal, most mutation-rich branch of the tree and, as a group, are more homogeneous than the lymph node metastases.

Patient C12’s tumor (Fig. 4B) partially resembled that of patient C58. Its phylogenetic tree also showed a group of lymph node metastases (L2, L3, L4) that either derived from the same ancestral clone or gave rise to each other, whereas other lymphatic lesions (L1) were seeded independently. Notably, as for patient C58, the closely related nodes also were in anatomical proximity. However, the patient’s liver metastasis (M1) did not arise from this subclone but instead had distinct origins in primary tumor area P8.

Another noteworthy case from the distinct origins group is patient C53 (Fig. 4C), who underwent resection of two metastases located in the right liver lobe and one metastasis in the left liver lobe. Phylogenetic reconstruction showed that the metastases in the right liver diverged relatively early. After their divergence, the primary tumor evolved further and independently gave rise to lymph node metastasis L1 and the left liver metastasis M3.

Further examples of cancers in the distinct origins category are shown in Fig. 4, D to F,



**Fig. 4. Phylogenetic trees of cancers with distinct origins of lymphatic and distant metastases.** (A to F) All trees except C12 (MSI case) (B) are drawn to scale and were constructed with the neighbor-joining method. Shading and clinical information is as noted for Fig. 3. r, right liver; l, left liver.

and fig. S4. In all these cases, the phylogenetic data indicate that lymph node metastases were not the source of distant metastases (also see explanatory schematic in fig. S10).

**Common clinicopathological variables do not correlate with origin classification**

Next, we examined whether our origin classification was correlated with (and thus potentially influenced by) any clinicopathological variables. We did not observe any significant differences in the number of positive nodes, the ratio of positive to examined nodes, the number of lymph nodes included in the final data set, the number of excluded nodes (fig. S11, A to D), the number of sampled primary tumor regions, the percentage of T3 versus T4 stage patients (no T1 or T2 stage tumors were part of this cohort), the distribution of primary tumor sizes, the presence of vascular invasion, or the fraction of patients with synchronous versus metachronous distant metastasis (fig. S12, A to E) between origin categories.

Most importantly, we found no association between origin and treatment history. Only one patient (C77) had neoadjuvant chemotherapy (table S3). In six patients, distant metastases were resected after the primary tumor and the lymph node metastases had already been removed, and all received treatment in the intervening time interval. Three of these patients (C69, C65, C36) fell into the common origins and three (C66, C39, C63) into the distinct origins category ( $P = 0.6$ ).

**Discussion**

The presence of lymph node metastases is an important prognostic factor for most cancers, but the underlying reason has been unclear. One prevailing model posits that lymph node metastases are precursors of distant metastases, and their surgical resection is necessary to attain a “cancer-free” state (34). An alternative model posits that distant metastases arise independently of lymph node metastases (16).

Our data show that lymph node metastases and distant metastases indeed often do have a

common origin. Although our phylogenies do not allow us to distinguish between sequential progression and common-subclonal origin, many phylogenies in the common origin category are compatible with seeding of distant metastases from lymph nodes.

However, in a majority of patients, we find strong evidence of independent origins of lymph node and distant lesions. If independent seeding is prevalent, what is the reason for the association of lymphatic and distant metastasis? It could be that the association is driven by the common origin subset of patients. An alternative possibility is that most cells in tumors belonging to the distinct origins category have the ability to metastasize. In such tumors, all cells that disseminate would have an increased likelihood of colonizing distant sites (35). Establishing lymph node metastases may be a more efficient process than establishing distant metastases and may therefore happen earlier and more frequently. This model would also be compatible with clinical observations, including the correlation between lymphatic and distant metastasis, the sometimes

modest benefits of lymphadenectomy, and the advantage of early primary resection (assuming that even in such highly metastatic cancers, the survival rate of disseminated cells is relatively low, so that the tumor needs to grow to a certain size in order to metastasize efficiently).

Most metastases in our cohort were resected from the liver, which is the most frequent distant site of colorectal metastasis (36). Because venous blood from the intestines reaches the liver directly through the portal vein, it is possible that liver metastases are preferentially seeded homogeneously. Cancer cells that migrate through lymph nodes enter the venous circulation in the subclavian vein. The first capillary bed that such cells encounter is the lung. It is therefore possible that lung metastases are more frequently seeded through the lymph nodes.

All cancers in our study were retrospectively collected specimens. Archival samples are mostly not suitable for whole-genome or exome sequencing because patient consent for such comprehensive genetic profiling was not obtained at the time of surgery. Conversely, polyguanine-repeat genotyping is a limited analysis of length polymorphisms in noncoding DNA. It does not produce any information about functional or disease-related genes. Raw data produced by our method contain the lengths of polymerase chain reaction (PCR) amplicons in arbitrary units, allowing for complete disclosure of mutation information while making patient identification impossible. Therefore, polyguanine-repeat analysis represents a safe and effective method for studying tumor evolution in a patient population that would otherwise be inaccessible.

We conclude that the evolutionary relationship between lymphatic and distant metastases can take on two different forms in colorectal cancer. In the future, it will be important to de-

termine whether cancers in the common and distinct origin categories exhibit different clinical behaviors.

#### REFERENCES AND NOTES

- S. D. Nathanson, *Cancer* **98**, 413–423 (2003).
- W. L. McGuire, *Breast Cancer Res. Treat.* **10**, 5–9 (1987).
- L. A. Gervasi et al., *J. Urol.* **142**, 332–336 (1989).
- T. Naruke, K. Suemasu, S. Ishikawa, *J. Thorac. Cardiovasc. Surg.* **76**, 832–839 (1978).
- G. J. Chang, M. A. Rodriguez-Bigas, J. M. Skibber, V. A. Moyer, *J. Natl. Cancer Inst.* **99**, 433–441 (2007).
- American Cancer Society, Treatment of colon cancer, by stage; [www.cancer.org/cancer/colon-rectal-cancer/treating/by-stage-colon.html](http://www.cancer.org/cancer/colon-rectal-cancer/treating/by-stage-colon.html).
- J. B. O'Connell, M. A. Maggard, C. Y. Ko, *J. Natl. Cancer Inst.* **96**, 1420–1425 (2004).
- S. Y. Wong, R. O. Hynes, *Cell Cycle* **5**, 812–817 (2006).
- R. Virchow, *Die krankhaften Geschwülste: Dreissig Vorlesungen gehalten während des Wintersemesters 1862–1893 an der Universität zu Berlin* (Verlag von August Hirschwald, Berlin, 1863).
- C. M. McBride, *South. Med. J.* **71**, 1331–1333 (1978).
- W. S. Halsted, *Ann. Surg.* **46**, 1–19 (1907).
- J. Sleeman, A. Schmid, W. Thiele, *Semin. Cancer Biol.* **19**, 285–297 (2009).
- R. A. Weinberg, *Carcinogenesis* **29**, 1092–1095 (2008).
- B. Moynihan, *Surg. Gynecol. Obstet.* **6**, 463–466 (1908).
- J. E. Gervasoni Jr., S. Sbayi, B. Cady, *Ann. Surg. Oncol.* **14**, 2443–2462 (2007).
- B. Cady, *Arch. Surg.* **119**, 1067–1072 (1984).
- J. Engel, R. T. Emery, D. Hölzel, *Cancer Metastasis Rev.* **31**, 235–246 (2012).
- W. Liu et al., *Nat. Med.* **15**, 559–565 (2009).
- A. McPherson et al., *Nat. Genet.* **48**, 758–767 (2016).
- W. J. Gibson et al., *Nat. Genet.* **48**, 848–855 (2016).
- M. Gerlinger et al., *N. Engl. J. Med.* **366**, 883–892 (2012).
- S. Yachida et al., *Nature* **467**, 1114–1117 (2010).
- J. G. Reiter et al., *Nat. Commun.* **8**, 14114 (2017).
- M. C. Haffner et al., *J. Clin. Invest.* **123**, 4918–4922 (2013).
- K. Naxerova et al., *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1889–E1898 (2014).
- J. C. Boyer et al., *Hum. Mol. Genet.* **11**, 707–713 (2002).
- S. J. Salipante, M. S. Horwitz, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5448–5453 (2006).
- D. Frumkin, A. Wasserstrom, S. Kaplan, U. Feige, E. Shapiro, *PLOS Comput. Biol.* **1**, e50 (2005).
- S. J. Salipante, J. M. Thompson, M. S. Horwitz, *Genetics* **178**, 967–977 (2008).
- D. M. Endres, J. E. Schindelin, *IEEE Trans. Inf. Theory* **49**, 1858–1860 (2003).
- N. Saitou, M. Nei, *Mol. Biol. Evol.* **4**, 406–425 (1987).
- J. J. Salk et al., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20871–20876 (2009).
- R. R. Sokal, *Univ. Kans. Sci. Bull.* **38**, 1409–1438 (1958).
- J. P. Sleeman, B. Cady, K. Pantel, *Clin. Exp. Metastasis* **29**, 737–746 (2012).
- B. Vogelstein, K. W. Kinzler, *N. Engl. J. Med.* **373**, 1895–1898 (2015).
- V. Patanaphan, O. M. Salazar, *South. Med. J.* **86**, 38–41 (1993).

#### ACKNOWLEDGMENTS

We thank M. Nahrendorf, F. Swirski, J. Gerold, and T. Padera for helpful comments and careful review of the manuscript and N. Sasaki and V. Sasselli for their help with organoid culture. This work was supported by the Department of Defense W81XWH-10-0016 (R.K.J.), W81XWH-12-1-0362 (S.J.E.), and W81XWH-15-1-0579 (K.N.); National Human Genome Research Institute U54 HG007963 (T.C.); National Cancer Institute P01-CA080124 (R.K.J.) and R35-CA197743 (R.K.J.); Francis Crick Institute FC001169 (C.S.); Austrian Science Fund J-3996 (J.G.R.); National Foundation for Cancer Research (R.K.J.); and Ludwig Center at Harvard (S.J.E. and R.K.J.). The Program for Evolutionary Dynamics is supported in part by a gift from B. Wu and E. Larson. K.N. conceived and designed the study and performed experiments. K.N. and J.G.R. analyzed data. E.B. and J.K.L. reviewed tissue specimens and clinical records. M.v.d.W., A.R., H.C., and C.S. provided DNA samples. K.N., J.G.R., E.B., J.K.L., T.C., C.S., M.A.N., S.J.E., and R.K.J. discussed results and strategy. R.K.J. supervised the study. K.N. wrote the manuscript, which was revised and approved by all authors. Raw polyguanine-profiling data, distance matrices, and phylogenetic trees can be downloaded from <https://steelelab.mgh.harvard.edu/lymphmet> and from [datadryad.org](http://datadryad.org) (<http://dx.doi.org/10.5061/dryad.vv53d>).

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/357/6346/55/suppl/DC1](http://www.sciencemag.org/content/357/6346/55/suppl/DC1)  
Materials and Methods  
Figs. S1 to S14  
Tables S1 to S3  
References

23 August 2016; resubmitted 16 February 2017  
Accepted 8 May 2017  
10.1126/science.aai8515

## Origins of lymphatic and distant metastases in human colorectal cancer

Kamila Naxerova, Johannes G. Reiter, Elena Brachtel, Jochen K. Lennerz, Marc van de Wetering, Andrew Rowan, Tianxi Cai, Hans Clevers, Charles Swanton, Martin A. Nowak, Stephen J. Elledge and Rakesh K. Jain

*Science* **357** (6346), 55-60.  
DOI: 10.1126/science.aai8515

### Metastases undergo reconstruction

Cancer cells from primary tumors can migrate to regional lymph nodes and distant organs. The prevailing model in oncology is that lymph node metastases give rise to distant metastases. This "sequential progression model" is the rationale for surgical removal of tumor-draining lymph nodes. Naxerova *et al.* used phylogenetic methods to reconstruct the evolutionary relationship of primary tumors, lymph node metastases, and distant metastases in 17 patients with colorectal cancer (see the Perspective by Markowitz). The sequential progression model applied to only one-third of the patients. In the other two-thirds, distant metastases and lymph node metastases originated from independent subclones within the primary tumor.

*Science*, this issue p. 55; see also p. 35

ARTICLE TOOLS	<a href="http://science.sciencemag.org/content/357/6346/55">http://science.sciencemag.org/content/357/6346/55</a>
SUPPLEMENTARY MATERIALS	<a href="http://science.sciencemag.org/content/suppl/2017/07/06/357.6346.55.DC1">http://science.sciencemag.org/content/suppl/2017/07/06/357.6346.55.DC1</a>
RELATED CONTENT	<a href="http://stm.sciencemag.org/content/scitransmed/9/397/eaan0026.full">http://stm.sciencemag.org/content/scitransmed/9/397/eaan0026.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/9/384/eaai8504.full">http://stm.sciencemag.org/content/scitransmed/9/384/eaai8504.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/8/364/364ra153.full">http://stm.sciencemag.org/content/scitransmed/8/364/364ra153.full</a> <a href="http://stm.sciencemag.org/content/scitransmed/8/361/361ra138.full">http://stm.sciencemag.org/content/scitransmed/8/361/361ra138.full</a>
REFERENCES	This article cites 34 articles, 9 of which you can access for free <a href="http://science.sciencemag.org/content/357/6346/55#BIBL">http://science.sciencemag.org/content/357/6346/55#BIBL</a>
PERMISSIONS	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

Use of this article is subject to the [Terms of Service](#)