Ancient hot and cold genes and chemotherapy resistance emergence

Amy Wu[a], Qiucen Zhang[b], Guillaume Lambert[c], Zayar Khin[d], Robert A. Gatenby[d], Hyunsung John Kim[e], Nader Pourmand[e], Kimberly Bussey[f], Paul C. W. Davies[g], James C. Sturm[a], and Robert H. Austin[h,1]

[a]Princeton Institute for the Science and Technology of Materials, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544; [b]Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [c]Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637; [d]Moffitt Cancer Center, Tampa, FL 33612; [e]Department of Bioengineering, University of California, Santa Cruz, CA 95064; [f]The Biodesign Institute, Arizona State University, Tempe, AZ 85287; [g]Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287; and [h]Department of Physics, Princeton University, Princeton, NJ 08544

Contributed by Robert H. Austin, June 26, 2015 (sent for review May 27, 2014; reviewed by Eshel Ben-Jacob[2] and Michael W. Deem)

We use a microfabricated ecology with a doxorubicin gradient and population fragmentation to produce a strong Darwinian selective pressure that drives forward the rapid emergence of doxorubicin resistance in multiple myeloma (MM) cancer cells. RNA sequencing of the resistant cells was used to examine (*i*) emergence of genes with high de novo substitution densities (i.e., hot genes) and (*ii*) genes never substituted (i.e., cold genes). The set of cold genes, which were 21% of the genes sequenced, were further winnowed down by examining excess expression levels. Both the most highly substituted genes and the most highly expressed never-substituted genes were biased in age toward the most ancient of genes. This would support the model that cancer represents a revision back to ancient forms of life adapted to high fitness under extreme stress, and suggests that these ancient genes may be targets for cancer therapy.

cancer | emergence | ancient genes | cold | hot

Multiple myeloma (MM), a hematologic cancer that develops in the bone marrow, is usually incurable because chemotherapy resistance emerges (1). The emergence of resistance may be largely due to the fact that bone marrow represents a very complex environment, due to the spatial heterogeneity of the bone marrow structure and the nonuniform distributions of nutrient, oxygen, and drug (during chemotherapeutic treatment) (2). Recent studies of the bone marrow represent an ideal ecology to be reproduced by microfluidic systems, with designed in vitro complex environments with a functional hematopoietic niche (3). Glucose gradient or chemotherapy gradients have been used to study the phenotypic progression of cancer in complex environments (4, 5); now we add the compartmentalization of small, possibly clonal, communities within the gradient. Just as rapid fixation of drug-resistant bacterial mutants in a metapopulation can occur in an environment with drug gradients and connected microhabitats (6–8), we demonstrate that an ecologically designed microenvironment, with drug gradients and connected microhabitats, can drive the rapid emergence of resistance in MM. We then use transcriptome sequencing of the far more complex (than in bacteria) genomic substitution patterns in the evolved MM cancer cells to address the question: What is the role of both substitutions and nonsubstitutions in the evolved genomes of the resistant cells in driving drug resistance?

In our previous work, we analyzed the 2D motions of metastatic breast cancer cells at the single-cellular level within a drug gradient without any local population bottlenecks (5). These experiments lasted for 72 h (at most, three generations), and there were no microhabitats within drug gradients, nor was any genomic analysis made. Here we designed connected microhabitats (hexagonal arrays) in the cell region to mimic the porous bone marrow structure, creating a metapopulation within the drug gradient with local fixation possibilities and invasion of more-fit mutants into higher-toxicity environments (6), and then analyzed the genomic changes that emerged in such a short time.

A critical part of this paper is the frequency of substitutions and nonsubstitutions in the resistant cells that emerge. Certainly, substitutions play a general role in the evolution of drug resistance and, therefore, targeted therapy could require a mutational event to provide cancer cells with a strategy around the therapy. However, we also know that duplication of key genes is common. This is a well-known evolutionary event that allows asexually reproducing species to maximize their adaptability and overcome Mueller's ratchet (9). Gene duplication can also be one of the causes of upregulation of proteins. We will examine this aspect in this paper.

Materials and Methods

Our device is composed of an array of hexagons, with small passageways connecting the six sides of the hexagons with adjacent ones. The array is a rectangular shape with two parallel channels maintaining the boundary concentrations of drug (Fig. 1). Our protocol was to first inoculate the device with cells without a drug gradient and incubate the cells for 24 h to ensure that the cells were alive and formed a uniform layer (Fig. 1*C*). Once this was achieved, a drug gradient was put across the culture chamber by turning on two syringes containing growth media alone and growth media containing doxorubicin (Fig. 1*D*). The gradient became stable within 30 min, and the drug concentration decreased linearly from the high side to the zero side (Fig. S1).

Emergence of doxorubicin resistance occurred on the time scale of days in wild-type parental MM with a high-end doxorubicin concentration of 20 nM [×5 the IC_{50} under continuous exposure (10)] maintained in the top channel

Significance

There are two broad components of information dynamics in cancer evolution. One involves permanent changes in which genes are subject to gain or loss-of-function substitutions. This is well established and the main focus of cancer research. The other component is the information in the human genome and preservation of that content. The cancer cell potentially has access to all of this and can upregulate or downregulate any number of strategies used for survival and proliferation during embryogenesis, development, and normal adaptation to environmental stresses. We suggest that nonsubstituted genes may be critical targets for chemotherapy; these nonmutated genes may be the most fundamental ones for preservation of cancer cell fitness, especially if their expression level changes.

Author contributions: A.W., Q.Z., G.L., R.A.G., N.P., P.C.W.D., J.C.S., and R.H.A. designed research; A.W. and R.H.A. performed research; A.W., Z.K., R.A.G., K.B., and R.H.A. contributed new reagents/analytic tools; A.W., Q.Z., H.J.K., N.P., K.B., P.C.W.D., J.C.S., and R.H.A. analyzed data; and A.W., G.L., R.A.G., K.B., P.C.W.D., J.C.S., and R.H.A. wrote the paper.

Reviewers: E.B.-J., Tel Aviv University; and M.W.D., Rice University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: austin@princeton.edu.

[2]Deceased June 5, 2015.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1512396112/-/DCSupplemental.

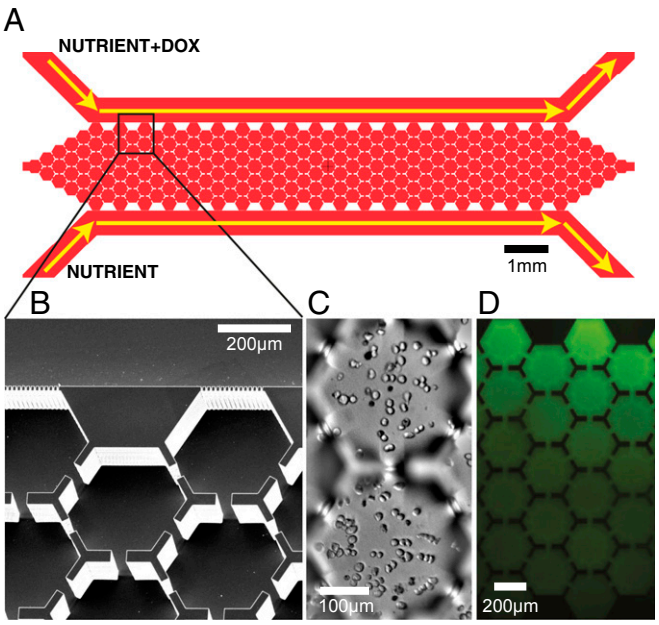

Fig. 1. Device layout and gradient characterization. (*A*) An overview of the entire microfluidic device, showing the flow of the nutrient streams and the nutrient + Doxorubicin (Dox) containing streams. The nutrient stream is growth medium, whereas the nutrient + Dox stream is growth medium + 20 nM Dox. (*B*) Scanning electron microscope image of the area of the array outlined by the box in *A*. (*C*) Image of MM cells in the device before imposing Dox gradient. (*D*) Image of the expected Dox concentration using the dye fluorescein as a marker. Prior work with similar structure to create a gradient (but without the walls to create microhabitats) gives a linear gradient (5).

(Dox+ in Fig. 2*A*). Despite the presence of doxorubicin, MM cells grew well and formed colonies initially near the nutrient channel (Dox− in Fig. 2*A*). Near the Dox+ channel, initially, MM growth was inhibited for 3 d, but resistant colonies ultimately appeared in a nonuniform manner across the gradient, as seen in Fig. 2*B*. The total increase in cell coverage during the experiment was only ×4 (from 15% to 60% as shown in Fig. 3*A*), indicating that, in the absence of cell death, only two generations of cells had passed before significant resistance had emerged. However, the increase in cell density was greatest at the midpoint of the gradient, where the doxorubicin concentration is still ×2 the IC_{50}, indicating the emergence of resistance across the gradient, and cell density proceeded toward the higher doxorubicin concentrations. Note that the overall population density only increases by a factor of ×4 in our 9-d experiments, which indicates that only four cell division cycles have occurred. Thus, the evolution of drug resistance in this experiment is relatively fast in terms of generation cycles.

We performed two control experiments to demonstrate the necessity of a drug gradient for the emergence of resistant MM cells. For experiment 1, we grew 10^6 cells in three tissue culture flasks with 10 nM of doxorubicin (replenished every 4 d). The choice of 10 nM is approximately driven by (*i*) ×2 the IC_{50} for this drug in a well-mixed tube and (*ii*) is midway in our gradient, where we see maximal growth in our device (see Fig. 2*B*). After 14 d, all of the cells lost viability based on trypan blue staining, demonstrating that the emergence of drug resistance cannot be achieved by a conventional single step of drug selection (11). For experiment 2, to confirm that the doxorubicin gradient instead of extracellular matrix induced the doxorubicin resistance in MM (12, 13), we pumped 10 nM of doxorubicin at both side channels of our devices, so that the concentration of drug was uniform at 10 nM throughout the microhabitat culture region, and compared the growth curves of MM with that in the same devices with doxorubicin gradient (0–20 nM per 2 mm) (Fig. 3*A*). The cells neither migrated nor grew after 3 d of uniform doxorubicin exposure in the devices. After 14 d, all cells in the devices (with or without gradients) were collected, and the viability was measured by trypan blue staining. The fact that all cells in the uniform doxorubicin environment lost viability indicates that the rapid emergence of doxorubicin resistance in MM only occurred in a gradient environment.

Approximately 10^4 MM cells were harvested after 14 d from the device [named as drug-resistant (DR) cells] and grown in a doxorubicin-free environment for 1 wk to expand the population size in the absence of stress. Then, the dose–response was performed to characterize resistance of DR cells versus the wild type (the parental MM cells, WT) (Fig. 3*B*) (*SI Text*). We found that the degree of cross-resistance (the IC_{50} of DR vs. WT) after 48 h of doxorubicin exposure increased 16-fold. This degree of cross-resistance required 10 mo to achieve by conventional protocols using step-wise increases of doxorubicin (10), indicating the more rapid ability of the MM cells to adapt to mutagenic stresses in a complex microenvironment, of profound impact to MM mortality in vivo.

Sampling MM cells from different regions within the drug gradient might unveil the degrees of resistance that emerged along the gradient. Ideally, spatially resolved sequencing along the drug gradient would allow construction of the resistance phylogeny trajectory. At this point, we grouped the cells from the gradient for dose–response characterization (Fig. 3*B*) and for sequencing analyses, so we have to leave where the resistance evolves along the gradient and the true role of chemotaxis as an open question.

RNA Sequencing Analysis of Resistant Cells

We performed RNA sequencing analysis of WT and DR samples to identify the expressed substitutions, single-nucleotide variants comparing with human reference genome (GRCh37), and differential expression levels of genes after the emergence of resistance. Note that the observed substitutions from sequencing data are the convolution of mutation and selection, although, in many cases of synonymous substitutions, there probably is no selection at work.

We sequenced four samples; two were samples that evolved in the drug gradient (D) and two were grown in the device without any drug (WT). Each sample was composed of 10^4 cells collected from three microfluidic devices running simultaneously under the same conditions. Although there is certainly a spatial dependence to the evolved genomes in the DR cells, in this preliminary paper, we grouped all of the cells on the gradient together. Table S1 gives the total number of reads, and Table S2 gives the total number of covered bases in two separate sequencing runs. Dataset S1 gives the expression levels of the genes (10,714 genes) that we detected based on the read abundance.

By no means were all of the roughly 20,000 genes in the human reference genome (Genome Reference Consortium GRCh37) successfully sequenced at high enough high exon coverage for single nucleotide variant (SNV) density analysis of the gene. Investigation of the substitution density that occurred in the subset of genes that we were confident in was done by comparison of the substitutions in the transcriptome of initial MM cells (WT) and the evolved resistant cancer MM cells (DR) to the human reference genome (Genome Reference Consortium GRCh37); this yielded the number of substitutions within a given gene in both the WT and the DR cells. With this > 80% cut, we were left with 785 genes where we could confidently compute the SNV density for both the incoming WT and the evolved DR cells. Assuming that 3,804 genes is a minimal set of genes to be expressed at a given time (14), we tested SNV in/out density for ∼20% of the expected fraction of expressed genes at any one time. Thus, this work presents only a snapshot of the full genomic changes occurring in the progression of resistance in our device.

Of the 785 genes where we could do SNV density analysis, we called SNVs that were not present in WT but are found in the evolved cell genome DR de novo substitutions. Fig. S2 summarizes the criteria for successfully sequenced genes. Dataset S2 lists the genomic coordinates of all de novo SNVs in the DR cells.

Only a relatively small fraction (on the order of 20%) of the substitutions were nonsynonymous; the majority were presumably neutral, carried along during evolution as "passenger substitutions." Table S3 presents the fraction of nonsynonymous substitutions.

Length matters in calculation of SNV of the density in genes that have genetic substitutions. The absolute number of substitutions in a given gene, that is, the numbers of SNVs per gene, is widely applied as a way to find putative drivers of adaptation (15). However, genes vary tremendously in length, ranging from hundreds to millions of bases in total (intron and exon) length as

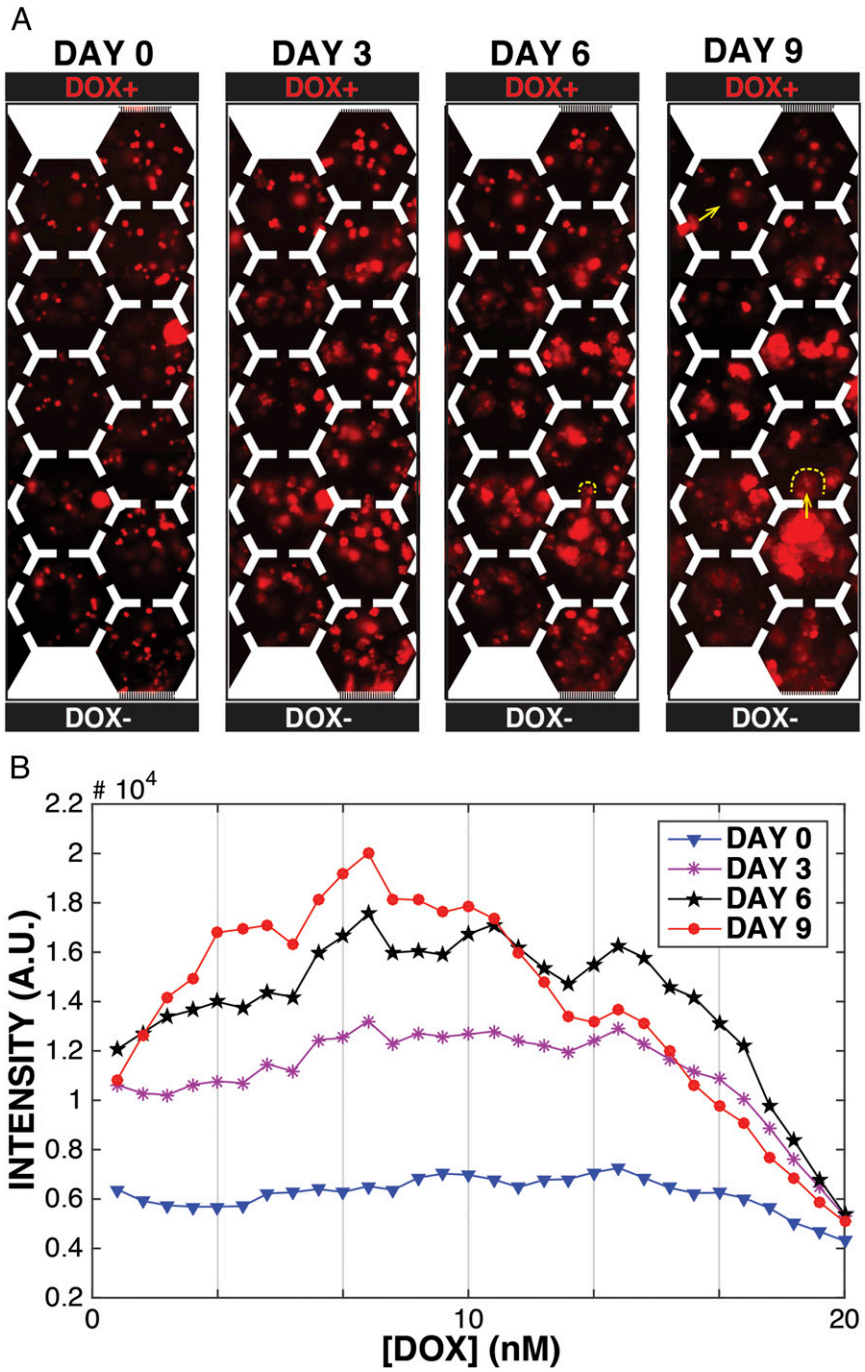

Fig. 2. Emergence of doxorubicin resistance of MM cells in a doxorubicin gradient. (*A*) Images of MM cells (8226/RFP) under a doxorubicin gradient (0–20 nM per 2 mm) in time series. Doxorubicin diffuses from the top to the bottom. Yellow dotted lines denote possible routes for MM cell migration. (*B*) Averaged density of MM cells (red fluorescence intensity) as a function of time and position in the doxorubicin gradient.

shown in Fig. S3, a histogram of the number of canonical human genes versus length.

Of course, if substitutions are random, then longer genes will show more substitutions than shorter genes; this does not mean that they are hot spots for substitutions, but rather that they are simply longer. The SNV density should not be a function of length in the random mutation model if there are no hot genes. For each gene(i) we are confident in assigning substitution densities, which we call the "per base substitution density" R_i as the number of de novo substitutions M_i divided by the length in base pairs of the successfully sequenced exon region (covered with > 20 reads) L_i. Because, at most, three substitutions were found in a gene, the binary nature of a substation in a given gene yields the nested curves shown in Fig. 4. We thus set the per gene substation density ρ (substitutions/bp) by dividing by the length L of the gene to correct for the smaller target size of short genes. Likewise, if we saw two substitutions, then the rate is $2/L$, etc. Fig. 4 shows a nested set, because the substitution rates are quantized at the gene scale, and a set of ascending curves as one approaches the origin (because of the $1/L$ effect by pure chance alone). However, note that as L decreases, one has many more genes. Averaging over the number of genes in a given window

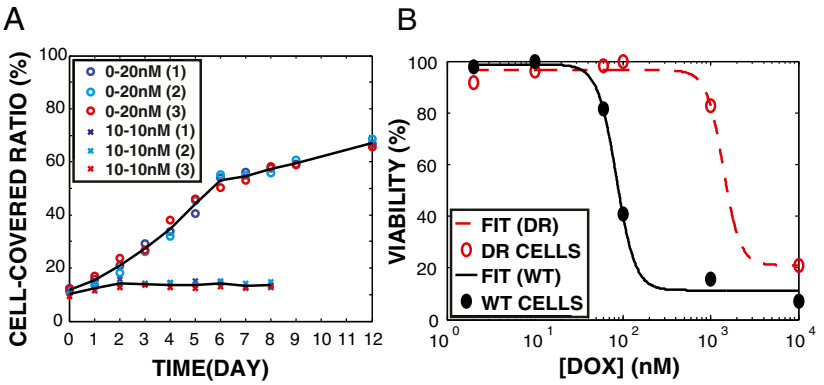

Fig. 3. Emergence of doxorubicin resistance in a doxorubicin gradient. (*A*) Growth curves of MM in microfluidic devices with doxorubicin: uniform (10 nM at both sides) vs. gradient (0–20 nM per 2 mm) doxorubicin exposure. Circles, three devices of gradient (0–20 nM per 2 mm) environment; crosses, three devices of uniform (10 nm) environment; lines, mean of three devices. (*B*) Doxorubicin dose–response (48-h exposure). Cells from drug gradient device (0–20 nM per 2 mm for 14 d, named as DR, red) vs. parental MM cells (WT, black). Data fitting was based on Hill equation (*SI Text*). The inhibitory concentration for 50% of control population (IC_{50}) of doxorubicin was increased 16-fold.

size, in our case 500 bp as shown in Fig. 4, gives a better representation of the density of substitution versus length. This process flattens the nested curves into a single curve, but there is still a tendency for more substitutions to occur in short genes compared with long genes.

The mean substitution density $\langle\mu\rangle$ is low enough that, even in the DR cells, most genes do not have substitutions, and hence presumably the substitutions per gene are governed by Poisson statistics (16). The power of the test is commonly set as 80% (17). Therefore, we followed the flowchart shown in Fig. S4 to determine successfully sequenced genes.

Margins of error in the per base substitution density on a given gene were determined by calculating the probability of the measured substitution density given the mean substitution density, assuming a binomial error distribution (*SI Text*). Genes with rates of substitutions, using the Bonferroni method (18), that were outside 95% confidence intervals (CI) from $\langle\mu\rangle$ we call hypermutated genes. Of the 785 successfully sequenced genes, 251 genes had at least one de novo SNV, and 163 genes were never mutated in DR nor WT samples, which we called cold genes. Cold genes represented 21% of the successfully sequenced genes, a sizable fraction.

Hot and Cold Genes

We call the genes with more than expected de novo nonsynonymous substitutions hot genes, and ignored synonymous replacements or nonsense replacements. We found 2,617 de novo SNVs, including 446 missense substitutions and 56 nonsense substitutions. In total, 439 genes had at least one de novo nonsynonymous SNVs. Among these genes, 45 genes were successfully sequenced (20*X* for 80% exon region).

We used the standard test for each successfully sequenced gene with de novo nonsynonymous SNV. Given a uniform probability for each position in a gene, a one-tailed binomial test was used to assess whether the observed substitution rate was significantly higher or lower than the binomial distribution. The mean substitution rate was calculated by the number of nonsynonymous SNVs divided by the total number of successfully sequenced bases ($502/13{,}714{,}589 = 3.7\times10^{-5}$). For each gene, the probability of detecting more substitutions than observed (*P* value) was calculated by the extreme upper tail binomial cumulative distribution function in Matlab.

Then, we performed multiple hypotheses tests using the standard Bonferroni procedure to look for significantly hypermutated genes. Given *P* values for 45 genes ($P_1, \ldots P_{45}$) to be tested at significance level $\alpha = 0.05$, we rejected null hypothesis (that the gene has expected number of substitutions) if $P_i < \frac{\alpha}{45} = 1.1\times10^{3}$. These significantly hypermutated genes are shown in Table 1.

These hot genes are involved in various biological functions such as apoptotic process, cell cycle, protein folding, cell division, metabolic process, oxidation–reduction process, and DNA topological change. Because the mechanism of action of doxorubicin includes topoisomerase II inhibition, DNA intercalation, and free radical generation, the abnormally high nonsynonymous substitution rates in these hot genes such as type 2 topoisomerase (TOP2B) may play a crucial role in elevated doxorubicin resistance in experiment.

In the subset of never-mutated (cold) genes, we restricted analysis to genes that had exceptionally high and low expression levels relative to the input WT cells, under the assumption that, among the large number of never-substituted genes, those with large changes in expression levels are playing a role in increase in the fitness of the resistant cells that emerged from our experiment. Because we did not do DNA sequencing, it was impossible to determine if expression-level changes relative to the input WT cells are due to copy number variations and/or changes in transcription factors. The relative expression levels of genes in DR vs. WT samples were based on the abundance of RNA reads (fragments per kilomegabases, FPKM) mapped to genes, as shown in Dataset S1 (*SI Text*). The log 2 ratio of DR to WT expression levels, $\log_2(\mathrm{FPKM_{DR}}/\mathrm{FPKM_{WT}})$, describes how much the expression levels changed after exposure to doxorubicin gradient. In other words, the greater (or less) the log 2 ratio for a given gene, the more it is upregulated (or downregulated) in DR samples. Histograms of numbers of all sequenced genes (with FPKM > 0.1 in both WT and DR samples) vs. log 2 ratio of expression levels are shown in Fig. S4.

The cold genes with extreme changes in expression levels seem to code for essential cellular functions, as shown in Tables 2 and 3. Upregulated PRDX4 is associated with spermatogenesis and oxidation–reduction process; upregulated PGK1 is associated with glycolysis and phosphorylation; downregulated PSMC1 is associated with cell cycle, DNA damage response, and apoptosis; downregulated ASB3 is associated with protein ubiquitination; downregulated RPS5 and RPLP0 are associated with translation. Although these zero-substitution genes shared by DR and WT cells do not directly explain drug resistance in DR samples, they could provide insights on protected regions of the genome during malignancy transformation.

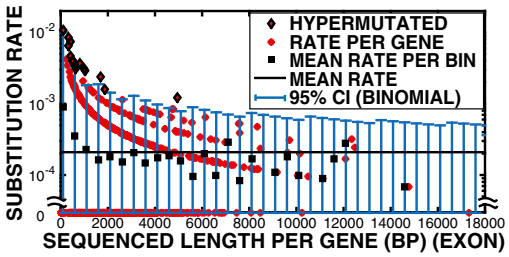

Fig. 4. Observed per base de novo substitution rate per gene vs. sequenced exonic length (bp) per gene. Red diamonds, genes that were successfully sequenced for more than 80% of exon region; black square, mean substitution density within a 500-bp window; black line, mean substitution density $\langle\mu\rangle = 2.0\times10^{-4}$/bp; blue error bars, 95% CI determined by binomial distribution (*SI Text*); black diamonds with red center, calls for hypermutated genes.

Table 1. Hot genes (nonsynonymous only)

Gene	SNVs	Exon bases	Sequenced exon bases (> 20×)	Substitution rate	Probability*	Age (×10^6 y)
HIST1H2:00 AM	2	487	430	4.7×10^{-3}	6.4×10^{-7}	unknown
SEH1L	2	3,506	2,922	6.8×10^{-4}	1.9×10^{-4}	1,530.3
CCAR1[†]	2	3,931	3,505	5.7×10^{-4}	3.2×10^{-4}	1,369.0
TUBA1A[‡]	1	978	978	1.0×10^{-3}	6.3×10^{-4}	2,269.5
ATAD2	2	6,017	4,892	4.1×10^{-4}	8.4×10^{-4}	3,556.3
MDH2[§]	1	1,305	1,207	8.3×10^{-4}	9.5×10^{-4}	2,535.8
TOP2B[¶]	2	5,422	5,197	3.8×10^{-4}	9.9×10^{-4}	3,556.3

*Probability to detect more substitutions than observed for this gene (*P* value).
[†]Biological process: apoptotic process, cell cycle.
[‡]Biological process: protein folding, microtubule-based process, cell division.
[§]Biological process: glucose metabolic process, oxidation–reduction process.
[¶]Biological process: DNA topological change, mitotic DNA integrity checkpoint.

The cold genes with large expression level changes are presumably important genes that cannot be substituted easily because they play a key role in fitness of the cells. It has been suggested that these cold genes might actually be very ancient genes representing a core functionality that cancer uses to maintain a basic fitness under high-stress conditions, as, presumably, early lifeforms must have experienced (19, 20).

It is possible to estimate roughly the age of genes by assessing the relative positions of the gene homologs in a phylogenetic tree. Such an analysis is shown in Fig. 5. We show there the histogram of calculated gene ages versus the human genome containing over 19,000 genes (solid black line) and the histogram of cold genes from our experiment. We found that the cold, zero-mutation genes we detected are older, on average, then all human genes, with the average age of 1.7 ± 1.0 billion years, compared with 19,786 human genes with an average age of 1.3 ± 0.9 billion years. The error bars, quite large, are simply a measure of the widths of the distributions and not a measure of the errors associated with this analysis, which are difficult to quantify at this stage. However, the large outliers of zero-mutation genes at 3.5 billion years of age are significant.

Discussion

We observed that, in a chemotherapy gradient landscape, resistance of MM cells to doxorubicin can emerge rapidly, as we expected from our bacterial work on antibiotic resistance emergence (8). We show that cold genes are more protected than many evolutionary conserved genes (21). We found that genes associated with nucleosome assembly and protein folding tend to mutate rapidly, whereas ancient genes associated with spermatogenesis, oxidation–reduction, and glycolysis are possibly protected and abnormally upregulated as a consequence of acquired drug resistance. The trade-offs may be associated with energy and time limitation in a harsh environment (22) and may be a strategy of cancer cells to evolve rapidly.

Because larger proteins have more surface area and more connections, it has been suggested that they encode more essential cellular functions than smaller proteins across various species (23). Also, ancient genes have been shown to evolve more slowly, and hence are cold, and express more "core" functional proteins compared with young genes (24, 25). As shown in Fig. 5, we observed the selection for the abnormally regulated ancient genes with slow evolution rate in emerged resistant cancer. Our sequencing results for emergence of resistant cancer address the integration on gene length, evolutionary rates, functional essentiality, and evolutionary ages; these properties have occurred in other species for guiding the animal body plans (25).

To be clear, we think there are two broad components of information dynamics in cancer evolution. One involves permanent changes in which genes are subject to gain or loss-of-function substitutions. This is well established and, unfortunately, the main focus of cancer research. The other component is the information in the human genome that is not mutated and in fact is protected from mutations. The cancer cell potentially has access to all of this and can upregulate or downregulate any number of strategies used for survival and proliferation during embryogenesis, development, and normal adaptation to environmental stresses. The concept that a mutation is needed to confer resistance is built into the Norton–Simon model, which has dictated cancer therapy practice for 5 decades (26).

The conventional view of the well-known emergence of drug resistance in cancer is that the initial stages are driven by substitutions that are random and independent events (27). In the conventional view, once a set of substitutions occur, selective pressure (i.e., chemotherapy for cancer) from the environment selects advantageous mutants out of this ensemble of substitutions, with a background of neutral fitness passenger substitutions carried along with the driver substitutions that change fitness (28). In the case of cancer cells under mutagenic stress, substitutions are perhaps random, but the frequency of the substitutions is increased by the stress-induced mutagenesis, so that drugs used in chemotherapy perversely can play a crucial role in the acceleration of the evolution of drug resistance (29). Here we chose the mutagenic chemotherapy drug doxorubicin as a mutagen, used a strong spatial selective pressure created by chemotherapy gradients, and used rapid fixation of substitutions within a metapopulation to accelerate the evolution of drug resistance in cancer, unlike the conventional protocol of gradually increasing in time the drug concentration (temporal drug gradients (11). Clearly, time-dependent chemotherapy gradients are also important in an in vivo setting, but that is beyond the scope of this paper.

An important question that we cannot answer here is: Were the de novo substitutions generated spontaneously or induced by stress? It is experimentally challenging to prove the existence of stress-induced mutagenesis, via upregulation of error-prone DNA polymerases or repressing DNA repair enzymes, because de novo mutation and selection usually come together (29). However, we observed upregulation of polymerase delta-interacting protein (POLDIP2) of DR cells in our experiments. POLDIP2 is known to support DNA polymerase λ in translesion synthesis, which often has

Table 2. Zero substitution, > 4× upregulated genes, ages

Gene	Exon bases	exon bases (> 20×)	age (×10^6 y)	$\log_2$(DR/WT)
PRDX4*	921	801	3,556.3	5.9
PGK1[†]	2,733	2,273	3,556.3	2.0
NOB1[‡]	1,775	1,535	2,269.5	5.9
CKS1B[§]	1,015	816	2,269.5	4.4

*Spermatogenesis, oxidation–reduction process.
[†]Gluconeogenesis, glycolysis, phosphorylation.
[‡]Proteosome biogenesis.
[§]Regulates MM growth.

Table 3. Zero substitution, > 4× downregulated genes, ages

Gene	Exon bases	exon bases (> 20×)	age (10^6 y)	$\log_2$(DR/WT)
PSMC1*	1,595	1,531	3,556.3	−3.7
ASB3†	1,275	1,094	3,556.3	−3.5
RPS5‡	741	728	3,556.3	−3.6
RPLP0§	1,304	1,124	3,556.3	−2.0

*Mitotic cell cycle, DNA damage response, apoptotic process.
†Protein ubiquitination, intracellular signal transduction.
‡Ribosomal protein component of 40S subunit.
§Ribosomal protein interacts with mitotic checkpoint protein MAD2B.

low fidelity (high propensity to insert wrong bases) on undamaged templates relative to regular polymerases and may induce de novo substitutions (29, 30). This error-prone recovery also protects DR cells from oxidative damage caused by doxorubicin (31, 32).

Another question is the role that cancer plays in development (25) and the transition from unicellular to multicellular behavior, and the role that cancer has played as an evolutionary variable (33). Because we show that up-expressed nonsubstituted genes and highly substituted genes are predominantly ancient genes, perhaps cancer represents a return to unicellularity that is represented by these crucial and ancient genes, with cancer allowing substitutions in, or abandoning higher-level genes associated with, multicellular cooperation (34). Clearly, with our limited data set, in this paper, we cannot address this question in a deeply quantitative way, but we hope we can point to different ways of viewing how cancer has influenced the process of development and its possible ancient origins.

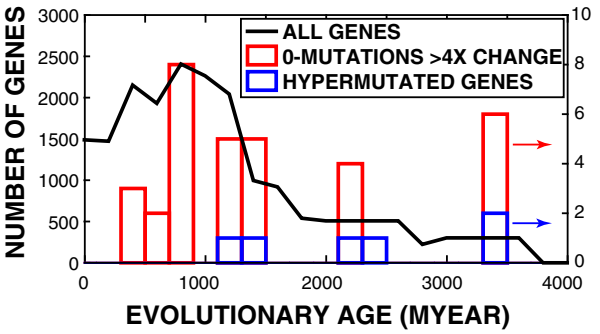

Fig. 5. Black is histogram of all human genes with age information (19786 in total). Blue is histogram of never-substituted genes with > 4× expression level changes (cold genes) vs. age. Red is histogram of hypersubstituted genes (nonsynonymous only, hot genes) vs. age.

ACKNOWLEDGMENTS. We thank our referees for provocative questions that greatly improved the paper, and Charles Lineweaver for discussions. The project described was supported by National Cancer Institute Grants U54CA143803 and U54CA143862.

1. Mahindra A, et al. (2012) Latest advances and current challenges in the treatment of multiple myeloma. *Nat Rev Clin Oncol* 9(3):135–143.
2. Mercier FE, Ragu C, Scadden DT (2012) The bone marrow at the crossroads of blood and immunity. *Nat Rev Immunol* 12(1):49–60.
3. Torisawa YS, et al. (2014) Bone marrow-on-a-chip replicates hematopoietic niche physiology in vitro. *Nat Methods* 11(6):663–669.
4. Liu L, et al. (2013) Minimization of thermodynamic costs in cancer cell invasion. *Proc Natl Acad Sci USA* 110(5):1686–1691.
5. Wu A, et al. (2013) Cell motility and drug gradients in the emergence of resistance to chemotherapy. *Proc Natl Acad Sci USA* 110(40):16103–16108.
6. Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc Int Congr Genet* 6:356–366.
7. Hermsen R, Hwa T (2010) Sources and sinks: A stochastic model of evolution in heterogeneous environments. *Phys Rev Lett* 105(24):248104.
8. Zhang Q, et al. (2011) Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science* 333(6050):1764–1767.
9. Muller HJ (1932) Some genetic aspects of sex. *Am Nat* 66(703):118–138.
10. Dalton WS, Durie BG, Alberts DS, Gerlach JH, Cress AE (1986) Characterization of a new drug-resistant human myeloma cell line that expresses P-glycoprotein. *Cancer Res* 46(10):5125–5130.
11. Calcagno AM, Ambudkar SV (2010) *Multi-Drug Resistance in Cancer*, Methods in Molecular Biology, ed Zhou J (Humana Press, Totowa, NJ), Vol 596, pp 77–93.
12. Damiano JS, Cress AE, Hazlehurst LA, Shtil AA, Dalton WS (1999) Cell adhesion mediated drug resistance (CAM-DR): Role of integrins and resistance to apoptosis in human myeloma cell lines. *Blood* 93(5):1658–1667.
13. Meads MB, Gatenby RA, Dalton WS (2009) Environment-mediated drug resistance: A major contributor to minimal residual disease. *Nat Rev Cancer* 9(9):665–674.
14. Eisenberg E, Levanon EY (2013) Human housekeeping genes, revisited. *Trends Genet* 29(10):569–574.
15. Lang GI, et al. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500(7464):571–574.
16. Chapman MA, et al. (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* 471(7339):467–472.
17. Colquhoun D (2014) An investigation of the false discovery rate and the misinterpretation of p-values. *R Soc Open Sci* 1(3):140216.
18. Shaffer JP (1995) Multiple Hypothesis Testing. *Annu Rev Psychol* 46:561–584.
19. Davies PCW, Lineweaver CH (2011) Cancer tumors as Metazoa 1.0: Tapping genes of ancient ancestors. *Phys Biol* 8(1):015001.
20. Davies P (2013) Exposing cancer's deep evolutionary roots. *Phys World* 26(7):37–40.
21. Bejerano G, et al. (2004) Ultraconserved elements in the human genome. *Science* 304(5675):1321–1325.
22. Aktipis CA, Boddy AM, Gatenby RA, Brown JS, Maley CC (2013) Life history trade-offs in cancer evolution. *Nat Rev Cancer* 13(12):883–892.
23. Tan T, Frenkel D, Gupta V, Deem MW (2005) Length, protein–protein interactions, and complexity. *Physica A* 350(1):52–62.
24. He J, Sun J, Deem MW (2009) Spontaneous emergence of modularity in a model of evolving individuals and in real networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 79(3 Pt 1):031907.
25. He J, Deem MW (2010) Hierarchical evolution of animal body plans. *Dev Biol* 337(1):157–161.
26. Simon R, Norton L (2006) The Norton-Simon hypothesis: Designing more effective and less toxic chemotherapeutic regimens. *Nat Clin Pract Oncol* 3(8):406–407.
27. Lawrence MS, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.
28. Bozic I, et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci USA* 107(43):18545–18550.
29. MacLean RC, Torres-Barceló C, Moxon R (2013) Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat Rev Genet* 14(3):221–227.
30. Hurley LH (2002) DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer* 2(3):188–200.
31. Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L (2004) Anthracyclines: Molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev* 56(2):185–229.
32. Maga G, et al. (2013) DNA polymerase δ-interacting protein 2 is a processivity factor for DNA polymerase λ during 8-oxo-7,8-dihydroguanine bypass. *Proc Natl Acad Sci USA* 110(47):18850–18855.
33. Sánchez Alvarado A (2012) Cellular hyperproliferation and cancer as evolutionary variables. *Curr Biol* 22(17):R772–R778.
34. Chen H, Lin F XH (2014) The degenerative evolution from multicellularity to unicellularity during cancer. arXiv:1408.3236v1.
35. Goutelle S, et al. (2008) The Hill equation: A review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol* 22(6):633–648.
36. Ding L, et al. (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481(7382):506–510.
37. Brody JP, Williams BA, Wold BJ, Quake SR (2002) Significance and statistical errors in the analysis of DNA microarray data. *Proc Natl Acad Sci USA* 99(20):12975–12978.
38. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
39. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
40. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
41. Futreal PA, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4(3): 177–183.