






In the format provided by the authors and unedited.

# Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination

Roby P. Bhattacharyya <sup>1,2,10</sup>, Nirmalya Bandyopadhyay<sup>1</sup>, Peijun Ma<sup>1</sup>, Sophie S. Son<sup>1</sup>, Jamin Liu <sup>1</sup>, Lorrie L. He<sup>1</sup>, Lidan Wu <sup>3</sup>, Rustem Khafizov<sup>3</sup>, Rich Boykin<sup>3</sup>, Gustavo C. Cerqueira<sup>1,4</sup>, Alejandro Pironti<sup>1</sup>, Robert F. Rudy <sup>1</sup>, Miles M. Patel<sup>1</sup>, Rui Yang<sup>1</sup>, Jennifer Skerry<sup>5</sup>, Elizabeth Nazarian<sup>6</sup>, Kimberly A. Musser<sup>6</sup>, Jill Taylor<sup>6</sup>, Virginia M. Pierce<sup>5</sup>, Ashlee M. Earl<sup>1</sup>, Lisa A. Cosimi<sup>7</sup>, Noam Shoresh<sup>1</sup>, Joseph Beechem<sup>3</sup>, Jonathan Livny<sup>1,10</sup> and Deborah T. Hung <sup>1,8,9,10\*</sup>

---

<sup>1</sup>Infectious Disease and Microbiome Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>2</sup>Infectious Diseases Division, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>3</sup>NanoString Technologies, Inc., Seattle, WA, USA. <sup>4</sup>Present address: Personal Genome Diagnostics, Ellicott City, MD, USA. <sup>5</sup>Microbiology Laboratory, Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. <sup>6</sup>Wadsworth Center, New York State Department of Health, Albany, NY, USA. <sup>7</sup>Infectious Diseases Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>8</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA. <sup>9</sup>Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA. <sup>10</sup>These authors jointly supervised this work: Roby P. Bhattacharyya, Jonathan Livny, Deborah T. Hung. \*e-mail: [hung@molbio.mgh.harvard.edu](mailto:hung@molbio.mgh.harvard.edu)

## Supplementary Information

### Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination

Roby P. Bhattacharyya<sup>+1,2</sup>, Nirmalya Bandyopadhyay<sup>1</sup>, Peijun Ma<sup>1</sup>, Sophie S. Son<sup>1</sup>, Jamin Liu<sup>1</sup>, Lorrie L. He<sup>1</sup>, Lidan Wu<sup>3</sup>, Rustem Khafizov<sup>3</sup>, Rich Boykin<sup>3</sup>, Gustavo C. Cerqueira<sup>1,4</sup>, Alejandro Pironti<sup>1</sup>, Robert F. Rudy<sup>1</sup>, Milesh M. Patel<sup>1</sup>, Rui Yang<sup>1</sup>, Jennifer Skerry<sup>5</sup>, Elizabeth Nazarian<sup>6</sup>, Kimberly A. Musser<sup>6</sup>, Jill Taylor<sup>6</sup>, Virginia M. Pierce<sup>5</sup>, Ashlee M. Earl<sup>1</sup>, Lisa A. Cosimi<sup>7</sup>, Noam Shores<sup>1</sup>, Joseph Beechem<sup>3</sup>, Jonathan Livny<sup>+1</sup>, Deborah T. Hung<sup>\*\*1,8,9</sup>

<sup>1</sup> Infectious Disease and Microbiome Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

<sup>2</sup> Infectious Diseases Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>3</sup> NanoString Technologies, Inc., Seattle, Washington, USA

<sup>4</sup> Present address: Personal Genome Diagnostics, Ellicott City, Maryland, USA

<sup>5</sup> Microbiology Laboratory, Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>6</sup> Wadsworth Center, New York State Department of Health, Albany, New York, USA

<sup>7</sup> Infectious Diseases Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA

<sup>8</sup> Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA

<sup>9</sup> Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA

\*Corresponding author: [hung@molbio.mgh.harvard.edu](mailto:hung@molbio.mgh.harvard.edu)

## Supplementary Methods

### *Strain diversity and assay generalizability:*

Genetic diversity within a species poses a fundamental challenge to the generalizability of bacterial molecular diagnostics, including transcription-based assays<sup>1</sup>. GoPhAST-R addresses this crucial challenge in a number of ways. First, for each pathogen-antibiotic pair, GoPhAST-R is trained and tested on a geographically and phylogenetically diverse set of strains: strains in this study were obtained from multiple geographic regions that sample across the entire phylogeny of each species (**Supplementary Fig 3**), notably including the CDC's Antibiotic Resistance Isolate Bank collection (<https://wwwn.cdc.gov/ARIsolateBank/>) that is intended as a test set for new diagnostic assays. Additionally, by targeting transcripts affected by antibiotics, which by definition affect core bacterial processes required for bacterial survival and whose transcriptional regulation is thus generally conserved<sup>1</sup>, GoPhAST-R measures responses that are also likely to be conserved and therefore generalizable.

One functional way to assess the generalizability of these antibiotic susceptibility signatures is to compare the optimal signatures chosen in the two phases of testing (**Supplementary Fig. 7**), and how well these chosen genes then performed on novel strains. Combining the initial derivation and validation cohorts from Phase 1 into a single, larger training cohort, we repeated feature selection and retrained the ensemble classifier on this larger set. The top 10 features chosen in this Phase 2 feature selection process were very similar to those chosen in Phase 1 (**Supplementary Table 3**), with 78% mean overlap in gene content, mean Jaccard similarity coefficient 0.67, and mean Spearman correlation coefficient 0.59 across all pathogen-antibiotic combinations. We then applied this refined classifier to predict susceptibility in a new test set of 25-30 isolates for each antibiotic (**Supplementary Fig. 8**). The fact that GoPhAST-R performed well on test strains that were selected randomly relative to training strains, that the sets of genes selected through iterative Phase 1 and 2 training were relatively similar, and that the same

classes of antibiotic elicit responses in similar pathways (**Supplementary Table 2**) and even homologous genes (**Supplementary Table 3**) across different species, all point to the ability of GoPhAST-R to account for the genetic diversity within a species.

Because diversity amongst clinical strains in gene content or sequence may hinder probe hybridization, we devised a homology masking algorithm to identify conserved regions of each target gene across all sequenced isolates from its species (see Methods and Supplementary Methods). By thus targeting the most conserved regions of core transcripts in the probe design process, GoPhAST-R also takes into account variability in genetic sequence of conserved genes in different strains. The phenotypic portion of the assay is particularly robust to sequence variation, both because it incorporates the behavior of multiple targets to provide redundancy, and because it measures fold-induction of the target gene by antibiotic, so a target gene that has mutated beyond recognition would not inform AST classification when registered as absent.

Ultimately, while we aimed to capture diversity in our initial sample set, testing of larger numbers of strains may be required. By employing a classification process built on machine-learning algorithms that can be iteratively refined as more strains are tested (**Supplementary Fig. 7**), GoPhAST-R is able to incorporate new diversity to asymptotically improve performance.

*Strategy for incorporating additional pathogen-antibiotic pairs:*

To extend GoPhAST-R to additional pathogen-antibiotic pairs beyond those defined here, the entire pathway described in this manuscript for signature derivation, from RNA-Seq to iterative phases of NanoString refinement and validation, will likely be required prior to implementation in a clinical setting. While some antibiotics elicit responses in predictable pathways, exemplified by fluoroquinolones up-regulating SOS-response transcripts, applying a novel diagnostic assay to a new pathogen-antibiotic pair demands rigor to meet clinical performance mandates. For

instance, when we applied this approach to *S. aureus* and *P. aeruginosa* treated with fluoroquinolones (**Supplementary Fig. 6**), we found that experimental derivation resulted in refined transcriptional signatures and control genes that were not predictable from prior assays on related pathogen-antibiotic pairs, often involving hypothetical or uncharacterized ORFs. The overall responses of both pathogens to fluoroquinolones included canonical DNA damage-responsive transcripts like *lexA*, *recA*, *recX*, *uvrA*, and *uvrB*, which were generally consistent with the genes identified for the other three gram-negative pathogens (**Supplementary Table 2**). Yet even for such a stereotypical response pathway, the specific genes selected from the RNA-Seq data that best distinguish susceptible from resistant isolates included features particular to each. In fact, *recA* was the only feature selected as a candidate responsive transcript in all five species; *lexA* and *uvrA* emerged in four of the five, but no other single transcript was selected in more than three, underscoring the importance of deriving each antibiotic response signature individually. In addition, derivation of control genes is critical for assay performance (see Supplemental Methods), and does not appear to be predictable across species or drug classes. This inability to predict the best-performing responsive and control genes by inference from other species highlights the need to individualize the expression signature for each pathogen-antibiotic pair, a process that is equivalent to the individualization currently required by CLSI to extend traditional AST assays to new pathogen-antibiotic pairs.

*Incorporating species identification into GoPhAST-R workflow:*

GoPhAST-R probes were designed to target regions conserved across all sequenced members of their parent species, thereby allowing each probeset to encode species identity in its reactivity profile. Since the NanoString platform can multiplex up to 800 probes in a single assay<sup>2</sup>, the actual deployed test would combine all 20 probes used for each pathogen-antibiotic pair (**Supplemental Table 3**) into a single multi-species probeset for each antibiotic, thereby providing simultaneous pathogen identification along with AST. Alternatively, species could be

identified prior to AST on the same NanoString platform using a more sensitive rRNA-based assay<sup>3</sup>.

*Proportionality of transcriptional response to antibiotic exposure:*

For each pathogen-antibiotic pair tested, the transcriptional profile of susceptible strains was distinct from that of resistant strains (**Supplementary Fig. 5a**), with the magnitude of the transcriptional response reflecting the MIC of the exposed isolate (**Supplementary Fig. 5b**). Sixth, the proportional relationship between transcriptional response and MIC underscores the biology that underpins the GoPhAST-R strategy: the more susceptible the strain, the greater its transcriptional response to antibiotic exposure. This relationship allows GoPhAST-R to be informed by established clinical breakpoint concentrations, thus leveraging decades of careful study linking *in vitro* strain behavior to clinical outcomes<sup>4</sup>. This relationship also explains why the majority of discrepancies between GoPhAST-R and broth microdilution occurred on strains with MICs close to the breakpoint. By contrast, the inability to map to MIC is considered a liability of genotypic assays, including WGS<sup>5</sup>.

*Carbapenemase detection in GoPhAST-R:*

By incorporating probes to simultaneously detect resistance determinants such as carbapenemase genes (**Fig. 2**), the genotypic component of GoPhAST-R can provide complementary evidence to reinforce its phenotypic call of resistance. This can be critical for the complex case of CRE detection<sup>6-11</sup>: even the American Type Culture Collection, the source of archived strain BAA2523, recognized this discrepancy in AST, noting that this carbapenemase-producing isolate was reported as carbapenem-susceptible upon deposition but tested resistant by other methods (<https://www.atcc.org/~ps/BAA-2523.ashx>). This is one of the isolates classified as resistant by GoPhAST-R but susceptible on standard MIC assays (**Supplementary Table 4**); it exhibits a dramatic inoculum effect (**Supplementary Fig. 1b-c**) that likely explains

this discrepancy, and clinical evidence is building that such carbapenemase-producing strains with large inoculum effects should be treated as resistant<sup>11-15</sup>.

As a hybridization-based assay, GoPhAST-R will tolerate mutation in its detection targets more robustly than PCR-based assays<sup>16,17</sup>. This enables GoPhAST-R to more readily detect resistance determinants with marked sequence variation, particularly exemplified for certain carbapenemase gene families. While the KPC and NDM gene families are quite conserved, conventional PCR-based detection of the IMP and VIM gene families has been challenging because of their genetic diversity<sup>18</sup> and the relative intolerance of PCR to point mutations in primer binding sites, especially towards the 3' end of the primer<sup>16,17</sup>. In contrast, hybridization is more tolerant to point mutations and is amenable to a multiplexed format that allows the inclusion of multiple probes to recognize different regions of the same target, and thus identify targets with greater diversity. For instance, GoPhAST-R includes 4 separate probe pairs to accommodate the diversity of the IMP family (**Supplementary Table S3**; see Supplemental Methods).

*Comparison to genomic AST prediction:*

GoPhAST-R offers a rapid alternative to current gold-standard phenotypic testing. One alternative approach under development is genotypic resistance testing, either through targeted genotypic assays such as nucleic acid amplification<sup>19-22</sup>, or through whole-genome sequencing. Whole genome sequencing (WGS) coupled with machine learning has promised the possibility of a more universal genomic approach to AST<sup>23-27</sup>, but many challenges remain, in terms of accuracy, generalizability, speed, cost, and deployability. While certain bacterial species or antibiotic classes are more amenable to genetic resistance prediction<sup>28,29</sup>, this approach remains far from generalizable<sup>5,30-32</sup>. Additionally, our inability to predict the emergence of new resistance mechanisms<sup>33</sup> means that genotypic resistance detection, whether targeted or

comprehensive, is fundamentally reactive as new resistance determinants are reported<sup>16,34-40</sup>. Genotypic resistance detection does, however, have the benefit of facilitating molecular epidemiology by allowing specific resistance mechanisms to be identified and tracked<sup>41,42</sup>. By integrating key aspects of both phenotypic and genotypic assays, GoPhAST-R realizes advantages of each approach.

### **Experimental and Analytical Details:**

#### *RNA extraction for RNA-Seq:*

After antibiotic treatment as described in main Methods section, cells were pelleted, resuspended in 0.5 mL Trizol reagent (ThermoFisher Scientific), transferred to 1.5 mL screw-cap tubes containing 0.25 mL of 0.1 mm diameter Zirconia/Silica beads (BioSpec Products), and lysed mechanically via bead-beating for 3-5 one-minute cycles on a Minibeadbeater-16 (BioSpec) or one 90-second cycle at 10 m/sec on a FastPrep (MP Bio). After addition of 0.1 mL chloroform, each sample tube was mixed thoroughly by inversion, incubated for 3 minutes at room temperature, and centrifuged at 12,000 xg for 15 minutes at 4°C. The aqueous phase was mixed with an equal volume of 100% ethanol, transferred to a Direct-zol spin plate (Zymo Research), and RNA was extracted according the Direct-zol protocol (Zymo Research).

#### *Library construction and RNA-Seq data generation:*

Illumina cDNA libraries were generated using a modified version of the RNAtag-Seq protocol<sup>43</sup>, RNAtag-Seq-TS, developed during the course of this study in which adapters are added to the 3' end of cDNAs by template switching<sup>44</sup> rather than by an overnight ligation, markedly decreasing the time, cost, and minimum input of library construction. Briefly, 250-500 ng of total RNA was fragmented, DNase treated to remove genomic DNA, dephosphorylated, and ligated to DNA adapters carrying 5'-AN<sub>8</sub>-3' barcodes of known sequence with a 5' phosphate and a 3' blocking group. Barcoded RNAs were pooled and depleted of rRNA using the RiboZero rRNA



depletion kit (Epicentre). Pools of barcoded RNAs were converted to Illumina cDNA libraries in 2 main steps: reverse transcription with template switching, then library amplification. RNA was reverse transcribed using a primer designed to the constant region of the barcoded adaptor with addition of an adapter to the 3' end of the cDNA by template switching using SMARTScribe (Clontech). Briefly, two primers are added to the reverse transcription reaction to facilitate template switching: primer AR2<sup>43</sup>, which primes SMARTScribe reverse transcriptase off of the ligated adapter, and primer 3Tr3<sup>43</sup>, which contains 3 protected G's at the 3' terminus to complement the C's added to the 3' end of newly synthesized cDNA by SMARTScribe and also contains a 5' blocking group to prevent multiple template-switching events. These primers were pre-incubated with rRNA-depleted, adapter-ligated RNA (at 8.33 uM of each primer) at 72°C x 3 min, then 42°C x 2 min, then added directly to a master mix containing SMARTScribe buffer (1x), DTT (2.5 mM), dNTPs (1mM each; NEB), SUPERase-In RNase inhibitor (1 unit; Invitrogen), and SMARTScribe reverse transcriptase enzyme (final primer concentration in reaction mixture: 5 uM each). This reaction mixture was incubated at 42°C x 60 min, then 70°C x 10 min, followed by addition of Exonuclease I (1 uL) and incubation at 37°C x 30 min. After 1.5x SPRI cleanup, the resulting cDNA library was PCR amplified using primers whose 5' ends target the constant regions of the ligated adapter (3' end of original RNA) and the template-switching oligo (5' end of original RNA) and whose termini contain the full Illumina P5 or P7 sequences. cDNA libraries were sequenced on the Illumina NextSeq 2500 or HiSeq 2000 platform to generate paired end reads.

#### *RNA-Seq data alignment:*

Sequencing reads from each sample in a pool were demultiplexed based on their associated barcode sequence. Barcode sequences were removed from the first read, as were terminal G's from the second read that may have been added by SMARTScribe during template switching. The resulting reads were aligned to reference sequences using BWA<sup>45</sup>, and read counts were

assigned to genes and other genomic features as described<sup>43</sup>. For each pathogen-antibiotic pair, a single reference genome was chosen for analysis of all four clinical isolates. This reference genome was selected by aligning a subset of RNA-Seq reads from each of the four isolates to all RefSeq genomes from that species and identifying the genome to which the highest percentage of reads aligned on average across all isolates. Since none of the isolates used for RNA-Seq have reference-quality genome assemblies themselves, and since four different isolates were used, not all genes in each isolate will be represented in the alignment. Yet for this application, any reads omitted due to the absence of a homologue in the reference genome used for alignment (i.e., accessory genes not shared by the reference) were assumed to be unlikely to be generalizable enough for diagnostic use. Using these criteria, the following reference genomes were chosen for alignment of RNA-Seq data for each of the following pathogen-antibiotic pairs: *K. pneumoniae* = NC\_016845 for meropenem and ciprofloxacin, and NC\_012731 for gentamicin; *E. coli* = NC\_020163 for meropenem, and NC\_008563 for ciprofloxacin and gentamicin; *A. baumannii* = NC\_021726 for meropenem, and NC\_017847 for ciprofloxacin and gentamicin. Note that for display purposes in **Figs. 1b, 3, and 4d**, and **Supplementary Figs. 6b, 8a, and 9a**, all responsive genes were named according to their homologues in the best-annotated reference available (NC\_016845 for *K. pneumoniae*, NC\_000913 for *E. coli*, and NC\_017847 for *A. baumannii*) in order to convey gene names that were as meaningful as possible, instead of simply gene identifiers. Read tables were generated, quality control metrics examined, and coverage plots from raw sequencing reads in the context of genome sequences and gene annotations were visualized using GenomeView<sup>46</sup>. Aligned bam files are deposited to the Sequence Read Archive (SRA) under study ID PRJNA518730.

#### *Selecting candidate responsive genes from RNA-Seq data:*

The DESeq2 package<sup>47</sup> was used to identify differentially expressed genes in treated vs untreated samples at each timepoint, in both susceptible and resistant strains. Analyses from

select timepoints are displayed as MA plots in **Fig 1a**, and from all timepoints in **Supplementary Fig. 2**. Since no statistically significant changes in transcription were observed in resistant strains, responsive gene selection was only carried out on susceptible isolates.

Complicating transcript selection is the fact that antibiotics interfere with growth of susceptible strains, resulting in the rapid divergence of culture density and growth phase of treated and untreated cultures, factors that alone affect the transcription of hundreds of genes that can mistakenly be interpreted as the direct result of antibiotic exposure but may not generalize across growth conditions. We thus expected that the resulting list of differentially expressed genes would represent both genes that respond primarily to antibiotic exposure, and genes that respond to ongoing growth that may be prevented by antibiotic treatment in susceptible strains, i.e. whose differential expression upon antibiotic exposure is more a secondary effect. As an example of this type secondary effect, consider a gene whose expression is repressed by increasing cell density, or nutrient depletion from the medium, as cells grow. In the presence of antibiotic, cells may never reach that cell density; therefore, this gene would exhibit higher expression in the antibiotic-treated culture (where it is not repressed) than in the untreated culture (where it is repressed). Without any correction, this gene would appear indistinguishable from one whose expression is induced by antibiotic, although this may be entirely a secondary effect. We reasoned that such “secondarily” regulated genes may be more dependent upon precise growth conditions (media type, temperature, cell density, cell state, etc – in other words, transcripts upregulated by progression towards stationary phase in minimal media will likely look different than that in rich media, etc), some of which may vary across clinical samples. By contrast, since antibiotics target core cellular processes, we hypothesized that the “direct” transcriptional response to antibiotic exposure would be more likely to be conserved across strains, which is critical for their success in a diagnostic assay. We therefore wished to focus on transcripts whose expression appeared to be a direct result of antibiotic exposure, rather than

this indirect result of the effects of an antibiotic on the progression of the strain to different culture densities.

To enrich for such genes specifically perturbed by antibiotic exposure, additional differential expression analyses were carried out using DESeq2 to identify genes whose expression varied in untreated samples over the timecourse of the experiment. Such genes were very common: >10% of the transcriptome was differentially regulated at some timepoints compared with others in our timecourses of *K. pneumoniae* and *E. coli* (though considerably fewer in *A. baumannii* cultures). We therefore imposed the additional requirement that any candidate responsive gene exhibit a greater degree of differential expression in time-matched antibiotic-treated vs untreated samples at  $\geq 1$  timepoint, than it did in any untreated timepoint – in other words, that antibiotics induce a degree of induction or repression that exceeds that which was achieved at any timepoint in the absence of antibiotics. To implement this, we imposed Fisher's combined probability test to combine p-values from each pairwise comparison, selecting those genes whose differential expression upon antibiotic treatment at a given timepoint exceeds their differential expression between any pair of points in the untreated timecourse, with adjusted p-value  $< 0.05$ . As an additional filter for gene selection, in order to be sure to target genes with sufficient abundance to be readily detected in the hybridization assay, only genes in the upper 50% of expression in each condition were considered.

For most pathogen-antibiotic pairs, this analysis resulted in the identification of hundreds of candidate antibiotic-responsive genes. We repeated this process (differential expression analysis + Fisher's method), using progressively higher thresholds for the fold-change threshold used in the statistical test for differential expression, by increasing the `lfcThreshold` parameter in DESeq2 (for all comparisons, i.e. antibiotic treatment and each pair of untreated timepoints used in Fisher's method) until the resulting list of candidate responsive genes was

60-100 long, the size we intended to target in phase 1 NanoString® assays. **Table S3** shows the fold-change thresholds used to generate the final candidate responsive transcript list for each pathogen-antibiotic pair.

This process was executed using custom scripts, available at

<https://github.com/broadinstitute/GeneSelection/> .

*Selecting candidate control genes from RNA-Seq data:*

To quantitatively compare the transcription of key antibiotic-responsive genes, it is critical to normalize for cell loading, lysis efficiency, and other experimental factors that may systematically affect absolute transcript abundance from a given sample. Such invariant transcripts (often referred to as “housekeeping” transcripts in qPCR) are crucial for scaling candidate responsive genes for comparison across samples, e.g. for comparing treated vs untreated samples. We therefore included control transcripts in our NanoString assay in order to normalize for these factors. We identified candidate control genes by seeking transcripts whose expression did not change in our RNA-Seq timecourses, either upon antibiotic treatment or with over the untreated timecourse. To find such genes, we imposed a statistical test to find transcripts whose expression did not change by more than a certain fold-change threshold in any of the treated or untreated samples by re-running DESeq2 using an inverted hypothesis test (`altHypothesis = “lessAbs”`), tuning the `lfcThreshold` parameter until we identified the 10-20 best control genes. **Supplemental Table 3** shows the fold-change thresholds used to generate the final candidate control transcript list for each pathogen-antibiotic pair.

*Gene Ontology (GO) term enrichment:*

For GO enrichment analysis, the same protocol was followed for responsive gene selection using DESeq2 and Fisher’s method (see “*Selecting candidate responsive genes from RNA-Seq*”).

*data*", above), with two exceptions. First, the `lfcThreshold` parameter was set to 0, in order to capture all differentially expressed genes. Second, genes of any expression level were considered, since sensitivity of detection was not a concern. This process produced a list of all genes that were differentially expressed upon antibiotic exposure to a greater extent than at any timepoint in the absence of antibiotic, over the full timecourse tested (0, 10, 30, and 60 min). These differentially expressed genes were named according to the reference genome that best matched the four strains used for RNA-Seq, as described (see "*RNA-Seq analysis*", above). GO terms were assigned to annotated genes from each reference genome by blasting the peptide sequences for each ORF from that reference genome against a local database of ~120 well-annotated reference strains from NCBI using blast2GO<sup>48</sup> (version 1.4.4). GO terms associated with the list of differentially expressed genes was then compared with all GO terms associated with the genome, and hypergeometric testing was deployed to identify GO terms that were enriched to a statistically significant extent among the differentially expressed genes, using the Benjamini-Hochberg correction for multiple hypothesis testing. A false discovery rate threshold of 0.05 was used to generate the list of enriched GO terms in **Supplemental Table 2**.

#### *Homology masking of selected responsive and control transcripts*

Within each candidate responsive or control gene, we identified regions of highest homology to target with NanoString probes. For each species, we compiled all complete reference genomes from RefSeq as of January 1, 2016, ran BLASTn to identify the closest homologue of each desired target from each reference genome, and eliminated targets without an annotated homologue in at least 80% of genomes. We then constructed a multi-sequence alignment and queried each sliding 100mer window to keep only those windows with at least one 100mer region of >97% nucleotide identity across all reference genomes; all sequences failing to meet this homology threshold were "masked", i.e., removed from consideration as targets for probe design. If no such region was found, the homology threshold was relaxed to >95% identity, then

to >92% identity; if no region with at least 92% identity was found, the transcript was deemed too variable to reliably target and thus eliminated from consideration entirely. The window size of 100 nucleotides was chosen because NanoString detection involves targeting with two ~50mer probes that bind consecutive regions<sup>2</sup>. The resulting homology-masked sequences, retaining only those regions of intended target transcripts with sufficient homology, were then provided to NanoString for their standard probe design algorithms<sup>2</sup>.

*Design of NanoString probes for carbapenemase and extended-spectrum beta-lactamase gene families:*

All gene sequences representing each targeted antibiotic resistance gene family (carbapenemases: KPC, NDM, OXA-48, IMP, VIM; ESBLs: CTX-M-15, OXA-10) were collected from representatives reported in three databases of antibiotic resistance genes: Resfinder<sup>49</sup>, ArDB<sup>50</sup>, and the Lahey Clinic catalog of beta-lactamases (<http://www.lahey.org/Studies>). Additional representatives of each family were identified by homology search (BLASTp, E-value < 10<sup>-10</sup>, >80% similarity) against the conceptual translation of genes identified in the genomes of isolates collected as part a multi-institute analysis of carbapenem-resistant *Enterobacteriaceae* specimens<sup>41</sup>. All other genes in the pan-genome of that cohort that did not meet the homology search criterion for inclusion as one of the targeted families were consolidated in an outgroup sequence database, which was used to screen for cross-reactivity. This outgroup contains many other non-targeted beta-lactamases, as well as the complete genomes of hundreds of *Enterobacteriaceae* isolates<sup>41</sup>. For each targeted antibiotic resistance gene family, target regions for NanoString probe design were identified as described above (see "*Homology masking of selected responsive and control transcripts*") by identifying regions with >95% sequence homology across 150 nucleotides in >90% of homologues within that family. In order to minimize risk of cross-reactivity with undesired targets, these conserved regions of the desired targets were then compared by BLASTn to the outgroup database, and any regions with

E-value < 10 were discarded. For the IMP gene family, no region of sufficient conservation could be identified due to sequence diversity within the family, consistent with reports that it is difficult to uniformly target by PCR<sup>18</sup>. We were able to identify four different regions that together were predicted to cover all IMP homologs from these databases, i.e., where each IMP homolog contained a stretch of sufficient homology to one or more of the four regions. We then submitted these regions to NanoString for probe design by their standard algorithms<sup>2</sup>, including four separate probe pairs for IMP (**Supplemental Table 3**). Signal from each of these four IMP probes was combined to yield a single combined total IMP signal (see “*NanoString data processing, normalization, and visualization*”, below).

*Lysate preparation for NanoString transcriptional profiling assays:*

Each strain to be tested was grown at 37°C in Mueller-Hinton broth to mid-logarithmic phase, and split into a treated sample, to which antibiotic was added at the CLSI breakpoint concentration, and an untreated control. Both samples were grown for the specified time (30-60 min), then a 100 uL aliquot of culture was added to 100 uL of RLT buffer (Qiagen) plus 1% beta-mercaptoethanol and mechanically lysed using either the MiniBeadBeater-16 (BioSpec) or the FastPrep (MP Biomedicals). This crude lysate was either used directly for hybridization, or frozen immediately and stored at -80°C, then thawed on ice prior to use.

*NanoString nCounter® assays:*

All Phase 1 and Phase 2 NanoString experiments (**Supplemental Fig. 7**) were performed on a NanoString nCounter® Sprint instrument, with hybridization conditions as per manufacturer’s recommendations, including a 10% final volume of crude lysate as input. Phase 1 experiments used probesets made with XT barcoded probe pools and were hybridized for 2 hours at 65°C, while Phase 2 experiments used probesets made with nCounter Elements™ probe pools plus cognate barcoded TagSets and were hybridized for 1 hour at 67°C, rather than the 16-24 hour



hybridizations as recommended by the manufacturer's protocol. Including 30-60 min for antibiotic exposure and these hybridizations, plus a 6 hour run for 12 samples, the total run time was under 8 hours for phase 2. Technical replicates for five strains run on separate days resulted in Pearson correlations of 0.95-0.99 for normalized data, consistent with expectations for this assay platform<sup>41</sup>, indicating that the shorter hybridization time did not affect reproducibility.

*Phylogenetic analysis of strains included in this study:*

The Genome Tree report was downloaded for each species from the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov>) in Newick file format and uploaded to the Interactive Tree of Life (iTOL; <https://itol.embl.de>)<sup>51</sup> for visualization and annotation. Strains from this study that were available on NCBI were identified using strain name or other identifying metadata from the NCBI Tree View file, cross-referencing the NCBI ftp server (<ftp://ftp.ncbi.nlm.nih.gov/pathogen/Results/>) as needed to confirm strain identity.

*Rapid transcriptional profiling with pilot NanoString Hyb & Seq™ assay platform*

For the rapid pilot GoPhAST-R experiment on a prototype Hyb & Seq™ instrument at NanoString (**Fig. 4**), we constructed pairs of capture probes (Probe A and Probe B) for all targets of interest such that each pair could uniquely bind to one target transcript. For Hyb & Seq chemistry (**Fig. 4b**), each Probe A contained a unique target binding region, a universal purification sequence, and an affinity tag for surface immobilization. Each Probe B contained another unique target binding region, a barcoded sequence for downstream signal detection, and a common purification sequence that is different from that of Probes A. For multiplexed RNA profiling, crude lysates were mixed with all capture and reporter probes in a single hybridization reaction and incubated on a thermocycler with heated lid at 65°C for 20 min. This

hybridization reaction enables formation of unique trimeric complexes between target mRNA, Probe A, and Probe B for each target.

We then performed three sequential steps of post-hybridization purification to ensure minimal background signal. Briefly, the hybridization product was first purified over magnetic beads coupled to oligonucleotides complementary to the universal sequence contained on every Probe B. The hybridization product was first incubated with the beads in 5x SSPE/60% formamide/0.1% Tween20 at room temperature for 10min in order to bind all target complexes containing Probes B, along with the free (un-hybridized) Probes B, onto the beads. Bead complexes were then washed with 0.1x SSPE/0.1% Tween20 to remove unbound oligos and complexes without Probes B. The washed beads were then incubated in 0.1x SSPE/0.1% Tween20 at 45°C for 10 min to elute the bound hybridized complexes off the beads. This second purification was carried out per manufacturer's instructions using Agencourt AMPure XP beads (Beckman Coulter) at a 1.8:1 volume ratio of beads to sample, in order to remove oligos shorter than 100 nt. This size-selective purification recovers the bigger hybridization complexes while removing smaller free capture Probes A and B. Eluates from these AMPure beads were purified over a third kind of magnetic beads coupled to oligonucleotides complementary to the common purification sequence contained on every Probe A, similar to the first bead purification, then eluted at 45°C. These triple-purified samples were driven through a microfluidic flow cell on a readout cartridge by hydrostatic pressure within 20 min. The flow cell was enclosed by a streptavidin-coated glass slide that can specifically bind to the affinity tag (biotin) of each Probe B, allowing the immobilization of purified complexes on the glass surface.

The cartridge with samples loaded was mounted on a Hyb & Seq prototype instrument equipped with an LED light source, an automated stage, and a fluorescent microscope. The barcoded region of each Probe A consisted of two short nucleic acid segments, each of which can bind to

one of ten available fluorescent bi-colored DNA reporter complexes as dictated by complementarity to the exact segment sequences. To detect each complex captured on the glass surface (**Fig. 4c**), photocleavable fluorescent color-coded reporters were grouped by their target segment location and introduced into the flow cell one pool at a time. Following each reporter pool introduction, the flow cell was washed with non-fluorescent imaging buffer to remove unbound reporter complexes and scanned by the automated Hyb & Seq prototype. Each field of view (FOV) was scanned at different excitation wavelengths (480, 545, 580 and 622 nm) to generate four images (one for each wavelength) and then exposed to UV (375nm) briefly to remove the fluorophore on surface-bound reporter probes by breaking a photocleavable linker. The flow cell was then subjected to a second round of probing with a new reporter pool targeting the second segment location on each Probe A. Thus, two rounds of probing, washing, imaging and cleavage completed one Hyb & Seq barcode readout cycle. In order to improve signal-to-noise ratio, 5 such cycles were completed for each assay. Between each cycle, the flow cell was incubated with low salt buffer (0.0033x SSPE/0.1% Tween20) to remove all bound reporter complexes without disrupting the ternary complex between Probe A, target mRNA, and Probe B.

A custom algorithm was implemented to process the raw images for each FOV on a FOV-by-FOV basis. This algorithm can identify fluorescent spots and register images between each wavelength and readout cycles. A valid feature is defined as a spot showing positive fluorescence readout for all barcoded segment locations in the same spatial position of each image after image registration. The molecular identity of each valid feature is determined by the permutation of color codes for individual rounds of barcode segment readout. In this implementation, the maximal degree of available multiplexing for a single assay using 10-plex reporter pools was  $10^2 = 100$  kinds for two-segment barcodes, but up to four-segment barcodes are available, permitting up to  $10^4 = 10,000$  distinct barcodes. This algorithm provides tabulated

results for the total raw count of each reporter barcode of interest identified in a single assay. These raw counts are used as input for subsequent data processing, visualization, and further analysis.

*NanoString data processing, normalization, and visualization:*

For each sample, read counts from each targeted transcript were extracted using nSolver Analysis Software (v4.070, NanoString, Seattle WA). Raw read counts underwent the following processing steps, all executed in R (version 3.3.3), utilizing the packages dplyr (version 0.7.4), xlsx (version 0.5.7), gplots (version 3.0.1), and DescTools (version 0.99.23):

1. Data aggregation: all data for a given pathogen-antibiotic pair, for a given phase of analysis (eg phase 1 or phase 2), was read in to a single data object so that all subsequent data processing steps were done together.
2. Positive control correction: per manufacturer's protocol, ERCC spike-ins were included in every hybridization at known concentrations, spanning the range of expected target RNA concentrations. For each sample, the geometric mean of counts from positive control probes targeting these ERCC spike-ins was calculated. This geometric mean was used to scale each remaining probe in the sample, in order to standardize across lanes for any systematic variation.
3. Negative control subtraction: per manufacturer's protocol, for each sample, the mean of negative control probes targeting ERCC spike-ins not present in the hybridization reaction were subtracted from the raw read counts for each target.
4. Failed probe removal: any control probe with fewer than 10 reads, or any responsive control with negative reads, after negative control subtraction in any sample was removed from all samples for a given pathogen-antibiotic pair, in order to omit transcripts whose content, sequence, or expression was too variable across strains.

5. Selection of optimal control probes: among the set of candidate control probes, across all strains in a given phase of analysis, the subset of these control probes that performed most consistently across samples was selected using a variation on the geNorm algorithm<sup>52</sup>. The principle behind this algorithm is that the per-cell expression of ideal control probes will not vary under any experimental conditions, and therefore, the ratio between expression levels of a set of ideal control probes will be constant (reflecting only the difference in cell number in each sample). We thus calculate the coefficient of variation of each control probe with the geometric mean of all control probes. In the ideal case, this coefficient of variation will be zero. The candidate control probe with the highest coefficient of variation is removed, and the process is repeated with the remaining control probes until the highest coefficient of variation is less than a threshold set to yield an acceptable number of non-operonic control transcripts, typically 4-8. For these experiments, this threshold was adjusted from 0.2 to 0.3 depending on the bacteria-antibiotic pair. Thresholds chosen, and the optimal control probes used at this threshold, are noted in **Supplementary Table 3**.
6. Control transcript normalization: the geometric mean of the optimal control probes was calculated for each sample and used to normalize all remaining read counts from that sample, i.e. for candidate responsive transcripts, and for carbapenemase or ESBL genes (if applicable), by dividing these corrected read counts by this geometric mean for each sample.
7. Calculation of fold-induction of normalized responsive transcripts by antibiotic: for each candidate responsive transcript, normalized counts from each antibiotic-treated strain were divided by normalized counts from untreated samples of the same strain. These fold-inductions of normalized expression for each candidate responsive transcript were used as input into machine learning algorithms, both reliefF for feature selection and the caret package for random forest classification.

8. Log-transformation of fold-induction data for responsive transcripts: for visualization, the natural logarithm of fold-inductions of normalized expression for each candidate responsive transcript was calculated and displayed using the heatmap.2 function of the gplots R package (version 3.0.1). For each set of strains,  $\ln(\text{fold induction})$  for each transcript was clustered using the default hclust function, and strains were ordered by MIC.
9. Combination of IMP probes: because of the variability of gene sequences in the IMP family, four separate IMP probes were designed, one or more of which was expected to recognize all sequenced members of this gene family. Following control gene normalization, signal from the four separate probes was added together to give a single IMP score.
10. Background subtraction for carbapenemase/ESBL gene detection: For each species, the subset of tested strains was identified for which whole-genome sequencing (WGS) data was available and none of the target beta-lactamase genes was found. From this subset, the arithmetic mean plus two standard deviations of the normalized signal from each probe (step 6) was calculated, and this mean + two standard deviations was subtracted from the normalized signal from each probe across all tested samples. All carbapenemases identified by WGS were detected above background, though the two *A. baumannii* isolates expressing bla<sub>NDM</sub> were only detected at very low levels. Background-subtracted data were log-transformed for visualization (any probe with a negative value after background-subtraction was set to 0.1 normalized counts for all standard nCounter experiments, or to 0.25 normalized counts for Hyb & Seq experiments, prior to log-transformation).

*One-dimensional projection of transcriptional data via squared projected distance (SPD) metric:*

Normalized, log-transformed fold-induction data from the ~60-100 responsive were collapsed into a one-dimensional projection that we call squared projected distance (SPD), essentially as described<sup>53</sup>. Conceptually, the transcriptional response of a test strain is placed on a vector in N-dimensional transcriptional space (where N = number of responsive genes, here ~60-100 per probeset) between the average position (i.e. centroid in transcriptional space) of a derivation set of susceptible strains (defined as SPD = 0) and the average position of a derivation set of resistant strains (defined as SPD = 1). All vector math was performed exactly as described<sup>53</sup> and implemented in R (version 3.3). For each pathogen-antibiotic pair, the same strains used for RNA-Seq were also used as the derivation set of two susceptible and two resistant strains, in order to ensure that the resulting projections of the remaining strains were not self-determined. In other words, only the strains used to select the transcripts to be used in the NanoString experiments (based on RNA-Seq) were used to set the average position of susceptible or resistant isolates; any tendency of other isolates to cluster at a similar SPD as these derivation strains, either susceptible or resistant, is thus due to a similarity in their transcriptional profiles. These derivation strains are labeled in **Supplementary Table 1** as “deriv\_S” and “deriv\_R” for susceptible and resistant strains, respectively. SPD data are plotted by CLSI class (**Supplementary Fig. 5a**) and by MIC (**Supplementary Fig. 5b**), showing a proportional relationship between MIC and this summative metric of transcriptional response to antibiotic exposure upon treatment at the breakpoint concentration (vertical dashed line).

*Approach to strain classification based on NanoString data:*

In order to select the most distinguishing features and to classify isolates as susceptible or resistant, we turned to machine learning algorithms, which we implemented in two phases (**Supplementary Fig. 7**).

In phase 1, NanoString XT probesets were designed targeting dozens (60-100) of antibiotic-responsive transcripts (**Supplementary Table 3**) selected from RNA-Seq data as described and used to quantify target gene expression from 18-24 isolates of varying susceptibility, both treated and untreated with the antibiotic in question, from which normalized fold-induction data for each responsive gene candidate was determined as described above. These isolates are partitioned into 50% training strains and 50% testing strains, randomly but informed by MIC: isolates are sorted in order of MIC and then alternately assigned to training and testing sets in order to ensure a balanced mix of isolates in each cohort across the full range of MICs represented by the strains in question. The only exceptions to random strain assignments to training vs testing sets in Phase 1 were: (1) intermediate isolates were not used for training, but were assigned to the validation cohort (and were grouped with resistant isolates for accuracy reporting, i.e., “not susceptible”), and (2) the two *E. coli* isolates with large meropenem inoculum effects were noted prior to randomization and deliberately assigned to the validation cohort, given the physiological basis for their discrepant transcriptional response from that of a conventional susceptible strain. From the training (derivation) cohort, the top 10 features were first selected using reliefF (see details below, “*Feature selection from NanoString data*”), then a random forest model was trained on this derivation cohort using the caret package, then implemented on the testing (validation) cohort, using only data from these top 10 selected features (see details below, “*Random forest classification of strains from NanoString data*”). Accuracy of GoPhAST-R in this phase was assessed by comparing predictions of the random forest model for the strains in the testing cohort, which it had never previously seen, with known susceptibility data for these strains (**Fig. 1c**, **Supplementary Fig. 2c**; **Supplementary Table 4**).

In phase 2, the training and testing cohorts from phase 1 were first combined into a single, larger training set, and selection of the top 10 responsive features were repeated using the same algorithms (reliefF). These represent our best-informed prediction of the 10 responsive



probes that most robustly discriminate between susceptible and resistant isolates, and are highlighted in **Supplementary Table 3** for each pathogen-antibiotic combination (column F = either “Phase 2” or “Top feature”). A new NanoString nCounter Elements probeset was then designed for each pathogen-antibiotic pair, targeting only these 10 transcripts as well as ~10 control probes that performed best in phase 1 (i.e. had the lowest coefficients of variation compared with the geometric mean of all control probes, using the variation on the geNorm algorithm described above; also indicated in **Supplementary Table 3**, column F). For *K. pneumoniae* + meropenem and ciprofloxacin, we proceeded to test an additional 25-30 strains using these focused phase 2 probesets, again quantifying target gene expression and normalized fold-induction of these responsive genes with and without antibiotic exposure. These data were supplied to the random forest classifier trained on all data from phase 1, and the resulting classifications of phase 2 strains were compared with known susceptibility data for these strains (**Supplementary Fig. 8b; Supplementary Table 4**). Of note, phase 2 deploys GoPhAST-R in exactly the way we envision it being deployed on true unknown samples: each of the phase 2 strains was an unknown, considered independently and not used at any point to train the model, only to assess its performance one strain at a time.

Every strain tested was an independent clinical isolate, with two minor exceptions. First, in the case of *A. baumannii* + ciprofloxacin, for which we did not have access to sufficient numbers of independent ciprofloxacin-susceptible *A. baumannii* isolates to train and test a classifier (only five out of 22 *A. baumannii* isolates). For this bacteria-antibiotic pair, we therefore ran biological replicates of the two susceptible strains used for RNA-Seq, RB197 (three replicates) and RB201 (two replicates). These biological replicates were grown from separate colonies in separate cultures, each split into treated and untreated samples. All three RB197 replicates ended up randomized to the phase 1 training set, while both RB201 replicates were randomized to the phase 1 testing set. Since we were not training on one biological replicate and testing on

another, the reported categorical agreement should not be confounded by excessive similarity between replicates. One additional linkage between isolates was that one *A. baumannii* isolate, RB197, exhibited two distinct colony morphotypes upon streaking onto LB plates: a dominant, larger morphotype, and a less abundant, smaller morphotype. The smaller morphotype was renamed RB197s and tested in both the meropenem and ciprofloxacin datasets, randomized to the testing (validation) cohort in both cases.

*Feature selection from NanoString data:*

For feature selection in both phase 1 and phase 2, we employed the reliefF algorithm<sup>54</sup> using the the CORElearn package (version 1.52.0) in R (version 3.3.3) to generate a list of features ranked in order of importance in distinguishing susceptible from resistant strains within the training set. The input to the reliefF algorithm was normalized fold-induction data from all responsive probes, and the CLSI classification, for each training isolate. (For this analysis, CLSI classification was simplified into two classes by grouping intermediate strains with resistant strains, in keeping with common clinical practice to avoid an antibiotic for which an isolate tests intermediate.)

The process by which reliefF generates its ranking is well-described elsewhere<sup>54</sup>. We chose the specific estimator algorithm (IEst parameter) “ReliefExpRank”, which considers the k nearest hits and misses, with the weight of each hit and miss exponentially decreasing with decreasing rank. We iterated five times (Itimes parameter = 5), with a separate 80% partition of the training data for each iteration, then averaged feature weight across each of these five iterations to generate the final ranked list. The output from this reliefF algorithm is a ranked list of features that best distinguish susceptible from resistant isolates; from this list, we chose to keep the top 10 features (featureCount parameter = 10). The same parameter values were chosen for feature selection for both phase 1 (i.e., on the half of the phase 1 data assigned to the training

set) and phase 2 (i.e., using all of the phase 1 data, for use in designing new probesets for *de novo* data acquisition in phase 2).

*Random forest classification of strains from NanoString data:*

To build a random forest classifier, we employed the caret (classification and regression training) package (version 6.0-78) in R (version 3.3.3) to classify strains in the testing cohort. Input data for this algorithm are normalized fold-inductions of the top 10 responsive genes selected by reliefF for both training and testing strains, and CLSI classifications for each training strain (again with intermediate and resistant isolates grouped together). This random forest model is a common example of an ensemble classifier<sup>55</sup> that embeds feature selection and weighting in building its models, which should mitigate risk for overtraining from including additional features from reliefF, since features not required for accurate classification need not be considered. It enacts 5-fold cross-validation on the training set, i.e. 80% sampling of the testing data, run 5 times, to optimize parameters including “mtry”, “min.node.size”, and “splitrule”, to build 500 trees (parameter “ntree” set to 500) based on prediction of the omitted training strains. After these hyperparameters are optimized through this cross-validation, an additional 500 trees are built using all of the training data and used to classify strains from the test set, one strain at a time. The resulting output is this classifier model that generates predictions for the classification of each test strain, reported as “probability of resistance” (probR) based on what fraction of trees ended up classifying the strain as resistant. (For instance, a strain with probR of 0.2 was classified as susceptible in 100 trees and as resistant in 400.) For quantitative assessment of accuracy, we used the prediction of the most likely class as the ultimate classification (i.e., if probR > 0.5, the classifier is predicting resistant; if probR < 0.5, the classifier is predicting susceptible). One might ultimately choose to set this threshold somewhere other than 0.5: since the cost of misclassifying a resistant isolate as susceptible (a “very major error” in the parlance of the FDA) is greater than the cost of misclassifying a

susceptible isolate as resistant, one might wish to label an isolate resistant if its probR is, say, 0.3. However, for simplicity, and to avoid overtraining on the relatively limited number of samples in this manuscript, we chose the default threshold of 0.5, accepting the classifier's prediction as to which state is more likely.

*Reproducibility of GoPhAST-R classification:*

Phase 2 probesets for meropenem susceptibility were combined with probes for carbapenemase and ESBL gene detection (**Supplementary Table 3**). For *K. pneumoniae* + meropenem, in addition to testing all phase 2 strains simultaneously for phenotypic AST and genotypic resistance determinants, we retested 23 of 24 phase 1 strains using the phase 2 probeset in order to capture their carbapenemase and ESBL gene content. This provides a set of effective technical replicates for assessing the robustness of our classifier, since all phase 2 genes are included as a subset of the phase 1 probeset, but all data were regenerated in a new NanoString experiment using the phase 2 probeset with added genotypic probes.

All 23 retested strains (11 susceptible, 12 resistant) were classified correctly based upon data from the phase 2 probeset; of these 23 strains, 12 (6 susceptible, 6 resistant) were phase 1 training strains (that were therefore not previously classified in phase 1), and 11 (5 susceptible, 6 resistant) were phase 1 testing strains that were classified the same way based upon data from the phase 2 probeset as they had been in phase 1 testing. The probability of resistance (probR) parameters for these 23 replicates from phase 1 (**Supplementary Table 4**) versus those from "re-classification" using data from the phase 2 probeset were highly correlated (Pearson correlation coefficient = 0.95). Note that because these same strains were used in training the random forest classifier, the results of re-classification of these retested strains are not included in the accuracy statistics reported elsewhere in this manuscript. The 100%

concordance observed for re-classification of these 23 strains is thus not a reflection of GoPhAST-R's accuracy, but does speak to its reproducibility.

*Blood culture processing:*

Under Partners IRB 2015P002215, 1 mL aliquots from blood cultures in the MGH clinical microbiology laboratory whose Gram stain demonstrated gram-negative rods were removed for processing. For simulated blood cultures, consistent with clinical microbiology laboratory protocol<sup>56</sup>, blood culture bottles were inoculated with individual isolates of each pathogen suspended in fetal bovine serum at <10 cfu/mL to simulate clinical samples and incubated in a BD BacTec FX instrument (BD Diagnostics; Sparks, MD) in the clinical microbiology laboratory at Massachusetts General Hospital, or on a rotating incubator at 37°C in our research laboratory at the Broad Institute. Once the BacTec instrument signaled positive (after 8.5-11.75 hours of growth), or after an equivalent time to reach the same culture density in the research laboratory (confirmed by enumeration of colony-forming units), 1 mL aliquots were removed for processing. Bacteria were isolated by differential centrifugation: 100 xg x 10 min to pellet RBCs, followed by 16,000 xg x 5 min to pellet bacteria. The resulting pellet was resuspended in 100 uL of Mueller-Hinton broth and immediately split into 5 x 20 uL aliquots for treatment with the indicated antibiotics (three antibiotics, plus two untreated samples, one for harvesting at 30 min to pair with the ciprofloxacin-treated aliquot and one at 60 min to pair with both meropenem- and gentamicin-treated aliquots). After the appropriate treatment time, 80 uL of RLT buffer + 1% beta-mercaptoethanol was added to 20 uL of treated bacterial sample, and lysis via bead-beating followed by NanoString detection were carried out as above (see "*Lysate preparation for NanoString transcriptional profiling assays*"). For real blood cultures, lysates were stored at -80C until organisms were identified in the laboratory by conventional means; only samples containing *E. coli* or *K. pneumoniae* were run on NanoString. GoPhAST-R results were compared with standard MIC testing in the MGH clinical microbiology laboratory, which were

also run on simulated cultures. Specimens were blinded until all data acquisition and analysis was complete. For head-to-head time trial compared with gold standard AST testing in the MGH clinical microbiology laboratory (subculture + VITEK-2), blood culture processing steps were timed in our research laboratory (Boston, MA, USA), then frozen and shipped to NanoString for transcript quantification on the prototype Hyb & Seq platform at NanoString (Seattle, WA, USA). A timer was restarted when lysates were thawed, and the total time at each site was combined to estimate the complete assay duration.

#### *Blood culture AST classification:*

Simulated blood cultures were classified using the same random forest approach as cultured strains, using the top 10 features selected during Phase 1 for each pathogen-antibiotic pair. This was implemented using leave-one-out cross-validation<sup>57</sup> rather than an even partitioning into training and testing because (1) feature selection was already complete, allowing multiple rounds of classifier training without requiring one unified model, and (2) given this, leave-one-out cross-validation (i.e., iteratively omit each strain once from training, test on the omitted strain, repeat with each strain omitted) allowed for training on the maximum number of strains.

#### **Supplementary References**

- 1 Wadsworth, C. B., Sater, M. R. A., Bhattacharyya, R. P. & Grad, Y. H. Impact of species diversity on the design of RNA-based diagnostics for antibiotic resistance in *Neisseria gonorrhoeae*. *Antimicrob Agents Chemother*, doi:10.1128/AAC.00549-19 (2019).
- 2 Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317-325, doi:10.1038/nbt1385 (2008).
- 3 Bhattacharyya, R. P. *et al.* Rapid identification and phylogenetic classification of diverse bacterial pathogens in a multiplexed hybridization assay targeting ribosomal RNA. *Sci Rep* **9**, 4516, doi:10.1038/s41598-019-40792-3 (2019).
- 4 CLSI. *Performance Standards for Antimicrobial Susceptibility Testing*. 28th edn, CLSI Supplement M100. Wayne, PA: Clinical and Laboratory Standards Institute (2018).
- 5 Ellington, M. J. *et al.* The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect* **23**, 2-22, doi:10.1016/j.cmi.2016.11.012 (2017).

- 6 Anderson, K. F. *et al.* Evaluation of methods to identify the *Klebsiella pneumoniae* carbapenemase in Enterobacteriaceae. *J Clin Microbiol* **45**, 2723-2725, doi:10.1128/JCM.00015-07 (2007).
- 7 Arnold, R. S. *et al.* Emergence of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *South Med J* **104**, 40-45, doi:10.1097/SMJ.0b013e3181fd7d5a (2011).
- 8 Control, C. f. D. & Prevention. Guidance for control of infections with carbapenem-resistant or carbapenemase-producing Enterobacteriaceae in acute care facilities. *MMWR Morb Mortal Wkly Rep* **58**, 256-260 (2009).
- 9 Gupta, V. *et al.* Phenotypic and genotypic characterization of carbapenem resistance mechanisms in *Klebsiella pneumoniae* from blood culture specimens: A study from North India. *J Lab Physicians* **10**, 125-129, doi:10.4103/JLP.JLP\_155\_16 (2018).
- 10 Nordmann, P., Cuzon, G. & Naas, T. The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *Lancet Infect Dis* **9**, 228-236, doi:10.1016/S1473-3099(09)70054-4 (2009).
- 11 Weisenberg, S. A., Morgan, D. J., Espinal-Witter, R. & Larone, D. H. Clinical outcomes of patients with *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae* after treatment with imipenem or meropenem. *Diagn Microbiol Infect Dis* **64**, 233-235, doi:10.1016/j.diagmicrobio.2009.02.004 (2009).
- 12 Adams-Sapper, S. *et al.* Rapid induction of high-level carbapenem resistance in heteroresistant KPC-producing *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* **59**, 3281-3289, doi:10.1128/AAC.05100-14 (2015).
- 13 Adler, A., Ben-Dalak, M., Chmelnitsky, I. & Carmeli, Y. Effect of Resistance Mechanisms on the Inoculum Effect of Carbapenem in *Klebsiella pneumoniae* Isolates with Borderline Carbapenem Resistance. *Antimicrob Agents Chemother* **59**, 5014-5017, doi:10.1128/AAC.00533-15 (2015).
- 14 Paterson, D. L. *et al.* Outcome of cephalosporin treatment for serious infections due to apparently susceptible organisms producing extended-spectrum beta-lactamases: implications for the clinical microbiology laboratory. *J Clin Microbiol* **39**, 2206-2212, doi:10.1128/JCM.39.6.2206-2212.2001 (2001).
- 15 Smith, K. P. & Kirby, J. E. The Inoculum Effect in the Era of Multidrug Resistance: Minor Differences in Inoculum Have Dramatic Effect on MIC Determination. *Antimicrob Agents Chemother* **62**, doi:10.1128/AAC.00433-18 (2018).
- 16 Paterson, G. K., Harrison, E. M. & Holmes, M. A. The emergence of *mecC* methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* **22**, 42-47, doi:10.1016/j.tim.2013.11.003 (2014).
- 17 Klungthong, C. *et al.* The impact of primer and probe-template mismatches on the sensitivity of pandemic influenza A/H1N1/2009 virus detection by real-time RT-PCR. *J Clin Virol* **48**, 91-95, doi:10.1016/j.jcv.2010.03.012 (2010).
- 18 Kaase, M., Szabados, F., Wassill, L. & Gatermann, S. G. Detection of carbapenemases in Enterobacteriaceae by a commercial multiplex PCR. *J Clin Microbiol* **50**, 3115-3118, doi:10.1128/JCM.00991-12 (2012).
- 19 Smith, M. *et al.* Rapid and accurate detection of carbapenemase genes in Enterobacteriaceae with the Cepheid Xpert Carba-R assay. *J Med Microbiol* **65**, 951-953, doi:10.1099/jmm.0.000310 (2016).
- 20 Sullivan, K. V. *et al.* Pediatric multicenter evaluation of the Verigene gram-negative blood culture test for rapid detection of inpatient bacteremia involving gram-negative organisms, extended-spectrum beta-lactamases, and carbapenemases. *J Clin Microbiol* **52**, 2416-2421, doi:10.1128/JCM.00737-14 (2014).
- 21 Traczewski, M. M., Carretto, E., Canton, R., Moore, N. M. & Carba, R. S. T. Multicenter Evaluation of the Xpert Carba-R Assay for Detection of Carbapenemase Genes in Gram-Negative Isolates. *J Clin Microbiol* **56**, doi:10.1128/JCM.00272-18 (2018).

- 22 Walker, T. *et al.* Clinical Impact of Laboratory Implementation of Verigene BC-GN  
Microarray-Based Assay for Detection of Gram-Negative Bacteria in Positive Blood  
Cultures. *J Clin Microbiol* **54**, 1789-1796, doi:10.1128/JCM.00376-16 (2016).
- 23 Allcock, R. J. N., Jennison, A. V. & Warrilow, D. Towards a Universal Molecular  
Microbiological Test. *J Clin Microbiol* **55**, 3175-3182, doi:10.1128/JCM.01155-17 (2017).
- 24 Bradley, P., den Bakker, H. C., Rocha, E. P. C., McVean, G. & Iqbal, Z. Ultrafast search  
of all deposited bacterial and viral genomic data. *Nat Biotechnol* **37**, 152-159,  
doi:10.1038/s41587-018-0010-1 (2019).
- 25 Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming  
clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* **13**, 601-612,  
doi:10.1038/nrg3226 (2012).
- 26 Li, Y. *et al.* Penicillin-Binding Protein Transpeptidase Signatures for Tracking and  
Predicting beta-Lactam Resistance Levels in *Streptococcus pneumoniae*. *MBio* **7**,  
doi:10.1128/mBio.00756-16 (2016).
- 27 Nguyen, M. *et al.* Developing an in silico minimum inhibitory concentration panel test for  
*Klebsiella pneumoniae*. *Sci Rep* **8**, 421, doi:10.1038/s41598-017-18972-w (2018).
- 28 Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for  
*Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* **6**, 10063,  
doi:10.1038/ncomms10063 (2015).
- 29 Consortium, C. R. *et al.* Prediction of Susceptibility to First-Line Tuberculosis Drugs by  
DNA Sequencing. *N Engl J Med* **379**, 1403-1415, doi:10.1056/NEJMoa1800474 (2018).
- 30 Bhattacharyya, R. P., Grad, Y. H. & Hung, D. T. in *Harrison's Principles of Internal  
Medicine* (eds J.L. Jameson *et al.*) Ch. 474, 3491-3504 (McGraw-Hill Education, 2018).
- 31 Rossen, J. W. A., Friedrich, A. W., Moran-Gilad, J., Genomic, E. S. G. f. & Molecular, D.  
Practical issues in implementing whole-genome-sequencing in routine diagnostic  
microbiology. *Clin Microbiol Infect* **24**, 355-360, doi:10.1016/j.cmi.2017.11.001 (2018).
- 32 Tagini, F. & Greub, G. Bacterial genome sequencing in clinical microbiology: a  
pathogen-oriented review. *Eur J Clin Microbiol Infect Dis* **36**, 2007-2020,  
doi:10.1007/s10096-017-3024-6 (2017).
- 33 MacLean, R. C. & San Millan, A. The evolution of antibiotic resistance. *Science* **365**,  
1082-1083, doi:10.1126/science.aax3879 (2019).
- 34 Caniaux, I., van Belkum, A., Zambardi, G., Poirel, L. & Gros, M. F. MCR: modern colistin  
resistance. *Eur J Clin Microbiol Infect Dis* **36**, 415-420, doi:10.1007/s10096-016-2846-y  
(2017).
- 35 Ford, B. A. mecC-Harboring Methicillin-Resistant *Staphylococcus aureus*: Hiding in Plain  
Sight. *J Clin Microbiol* **56**, doi:10.1128/JCM.01549-17 (2018).
- 36 Garcia-Alvarez, L. *et al.* Meticillin-resistant *Staphylococcus aureus* with a novel mecA  
homologue in human and bovine populations in the UK and Denmark: a descriptive  
study. *Lancet Infect Dis* **11**, 595-603, doi:10.1016/S1473-3099(11)70126-8 (2011).
- 37 Liakopoulos, A., Mevius, D. J., Olsen, B. & Bonnedahl, J. The colistin resistance mcr-1  
gene is going wild. *J Antimicrob Chemother* **71**, 2335-2336, doi:10.1093/jac/dkw262  
(2016).
- 38 Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in  
animals and human beings in China: a microbiological and molecular biological study.  
*Lancet Infect Dis* **16**, 161-168, doi:10.1016/S1473-3099(15)00424-7 (2016).
- 39 Ma, P., Laibinis, H. H., Ernst, C. M. & Hung, D. T. Carbapenem Resistance Caused by  
High-Level Expression of OXA-663 beta-Lactamase in an OmpK36-Deficient *Klebsiella  
pneumoniae* Clinical Isolate. *Antimicrob Agents Chemother* **62**, doi:10.1128/AAC.01281-  
18 (2018).
- 40 Sun, J., Zhang, H., Liu, Y. H. & Feng, Y. Towards Understanding MCR-like Colistin  
Resistance. *Trends Microbiol* **26**, 794-808, doi:10.1016/j.tim.2018.02.006 (2018).



- 41 Cerqueira, G. C. *et al.* Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. *Proc Natl Acad Sci U S A* **114**, 1135-1140, doi:10.1073/pnas.1616248114 (2017).
- 42 Woodworth, K. R. *et al.* Vital Signs: Containment of Novel Multidrug-Resistant Organisms and Resistance Mechanisms - United States, 2006-2017. *MMWR Morb Mortal Wkly Rep* **67**, 396-401, doi:10.15585/mmwr.mm6713e1 (2018).
- 43 Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat Methods* **12**, 323-325, doi:10.1038/nmeth.3313 (2015).
- 44 Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892-897, doi:10.2144/01304pf02 (2001).
- 45 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 46 Abeel, T., Van Parys, T., Saeys, Y., Galagan, J. & Van de Peer, Y. GenomeView: a next-generation genome browser. *Nucleic Acids Res* **40**, e12, doi:10.1093/nar/gkr995 (2012).
- 47 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 48 Gotz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**, 3420-3435, doi:10.1093/nar/gkn176 (2008).
- 49 Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67**, 2640-2644, doi:10.1093/jac/dks261 (2012).
- 50 Liu, B. & Pop, M. ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res* **37**, D443-447, doi:10.1093/nar/gkn656 (2009).
- 51 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).
- 52 Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).
- 53 Barczak, A. K. *et al.* RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities. *Proc Natl Acad Sci U S A* **109**, 6217-6222, doi:10.1073/pnas.1119540109 (2012).
- 54 Robnik-Šikonja, M. & Kononenko, I. Theoretical and Empirical Analysis of Relief and RRelief. *Machine Learning* **53**, 23-69, doi:10.1023/a:1025667309714 (2003).
- 55 Liaw, A. & Wiener, M. *Classification and Regression by RandomForest*. Vol. 23 (2001).
- 56 Clark, R. B., Lewinski, M. A., Loeffelholz, M. J. & Tibbetts, R. J. *Cumitech 31A, Verification and validation of procedures in the clinical microbiology laboratory*. (ASM Press, 2009).
- 57 Efron, B. & Gong, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *American Statistician* **37**, 36-48, doi:Doi 10.2307/2685844 (1983).