differentiations, even for cells from the same individual. To characterize this resulting cellular heterogeneity, they sequenced RNA from individual hiPSC-derived sensory neurons from one donor, demonstrating that, while 63% of cells formed a tight cluster expressing sensory neuronal genes, the remaining 37% of cells expressed genes more typical of fibroblasts. Using computational deconvolution, they found that the level of this fibroblast-like signature varied across hiPSC-derived neuronal cell lines in their data, even between multiple differentiations from the same hiPSC line. This is consistent with our own findings that variation in cell type composition between neuronal differentiations is driven by a fibroblast-like signature[14]. Moreover, they show that genes that are significantly upregulated following neuronal differentiation from hiPSCs, including ones critical to neuronal function, were the most variable. This high expression variability in hiPSC-derived neurons, which are not fully identical to their in vivo counterparts, is an important caveat that limits the power of these models.

hiPSC-based eQTL analyses

Despite the limitations to hiPSC-based studies more clearly delineated here, Schwartzentruber et al.[1] successfully applied allele-specific methods to map 1,403 expression quantitative trait loci (eQTLs) and 6,318 chromatin accessibility QTLs (caQTLs) at a false discovery rate (FDR) of 10%. Here, as in the NextGen eQTL studies[5,6], hiPSC-based eQTL analyses confirmed in vivo findings reported in GTEx, but also discovered novel eQTLs missed by tissue-level analyses. These positive eQTL findings reflect the larger effect sizes relative to other variables of interest, such as disease status or electrophysiological properties.

On the basis of the degree of expression variation observed in this dataset, Schwartzentruber et al.[1] estimate that recall-by-genotype studies using hiPSC-derived neurons will require at least 20–80 unrelated individuals to detect the effects of regulatory variants with even moderately large effect sizes. This is a helpful insight, emphasizing the necessity of further increasing the overall size of hiPSC-based studies of complex genetic disease, even at the cost of eliminating replicate hiPSC clones for any given individual[14,15]. Moreover, it supports an urgent need to develop isogenic models to query the functional impact of common variants, many of which are not conserved in rodents and must be studied in human cells.

Overall, these findings raise concerns about cell type heterogeneity and the associated expression variation across multiple differentiations that currently limit the power of the hiPSC platform. This raises an important challenge for the field when selecting differentiation or induction protocols for hiPSC-based molecular and cellular analyses (Fig. 1). Current differentiation protocols tend to be evaluated on the basis of cellular yield (i.e., the percentage of cells positive for one or more cell-type-specific markers) across a handful of hiPSC lines. Moving forward, it will be critical to assess expression variance between differentiations using cells from the same and different individuals, to test the extent that the donor effect is conserved. The power of hiPSC-based models for molecular and phenotypic studies of disease risk depends on this retention of the donor-specific component of gene expression. ❐

Gabriel E. Hoffman[1,2*] and Kristen J. Brennand[1,2,3,4,5*]

[1]*Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.* [2]*Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.* [3]*Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.* [4]*Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA.* [5]*Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.*

**e-mail: gabriel.hoffman@mssm.edu; kristen.brennand@mssm.edu*

Published online: 22 December 2017
https://doi.org/10.1038/s41588-017-0017-4

References

1. Schwartzentruber, J. et al. *Nat. Genet.* https://doi.org/10.1038/s41588-017-0005-8 (2017).
2. GTEx Consortium. *Nature* **550**, 204–213 (2017).
3. Fromer, M. et al. *Nat. Neurosci.* **19**, 1442–1453 (2016).
4. D'Antonio, M. et al. *Stem Cell Rep.* **8**, 1101–1111 (2017).
5. Carcamo-Orive, I. et al. *Cell Stem Cell* **20**, 518–532 (2017).
6. DeBoever, C. et al. *Cell Stem Cell* **20**, 533–546 (2017).
7. Panopoulos, A. D. et al. *Cell Stem Cell* **20**, 505–517 (2017).
8. Warren, C. R. et al. *Cell Stem Cell* **20**, 547–557 (2017).
9. Pashos, E. E. et al. *Cell Stem Cell* **20**, 558–570 (2017).
10. Panopoulos, A. D. et al. *Stem Cell Rep.* **8**, 1086–1100 (2017).
11. Banovich, N.E. et al. Preprint at *bioRxiv* https://doi.org/10.1101/091660 (2016).
12. Rouhani, F. et al. *PLoS Genet.* **10**, e1004432 (2014).
13. Kilpinen, H. et al. *Nature* **546**, 370–375 (2017).
14. Hoffman, G.E. et al. Preprint at *bioRxiv* https://doi.org/10.1101/185546 (2017).
15. Germain, P. L. & Testa, G. *Stem Cell Rep.* **8**, 1784–1796 (2017).

Competing interests

The authors declare no competing financial interests.

PLANT MICROBIOME

Bacterial genomics of plant adaptation

What allows bacteria, both pathogens and mutualists alike, to survive in close association with a eukaryotic host? A new study performed a large-scale comparative genomics analysis to identify novel genetic and genomic traits that are enriched in plant-associated bacterial taxa.

Ryan A. Melnyk and Cara H. Haney

Bacteria are found in every conceivable niche on the planet. The ability to associate with either autotrophic (i.e., plant) or heterotrophic (i.e., animal) eukaryotes provides bacteria with access to otherwise unavailable carbon sources[1]. Bacteria that fill host-associated niches can be mutualistic, pathogenic or neither. Regardless of the effect on the host, for bacteria to live in relatively nutrient-rich host-associated environments, they must have the genetic potential to utilize host nutrients, avoid or suppress host immunity, and compete with the other microbes that are vying for the same niches[2]. Identifying the genetic arsenal that allows bacteria to compete in host-associated niches including the plant root ('rhizosphere')

and shoot ('phyllosphere') is essential for engineering or manipulating microbial communities for agricultural improvement and for combatting emergent pathogens. In a new study, Asaf Levy, Isai Salas Gonzalez and colleagues[3] describe a large-scale comparative genomics approach to identify bacterial genes that are enriched in the genomes of plant-associated bacteria.

Genomics of plant association

There is an ever-growing collection of high-quality publicly available bacterial genome sequences from plant, animal and environmental isolates. Comparative genomics has been successful in identifying bacterial genes that correlate with specific functional outcomes within a single genus[4–7]. Levy et al.[3] used nearly 4,000 high-quality, non-redundant bacterial genomes spanning multiple bacterial phyla and plant origins, including 377 newly sequenced genomes from rhizosphere isolates from the reference plant *Arabidopsis thaliana*, a model tree (*Populus* spp.) and corn (*Zea mays*). They then used a comparative genomics approach independent of host, bacterial taxa or predicted function, to identify gene families that are enriched in plant-associated bacterial taxa relative to environmental or animal-associated strains. Levy et al.[3] successfully identified homologs of known bacterial genes involved in colonization, pathogenesis or provision of nutrients to plants (Fig. 1). These include nodulation genes in rhizobia species, nitrogen fixation genes in *Burkholderia* and type 3 secretion system genes in *Xanthomonas*, a genus that includes plant pathogens. Enrichment of gene families known to be important for plant–microbe interactions confirms that their large-scale approach was successful in identifying genes that are important for plant association.

Beyond the usual suspects in plant–microbe interactions, Levy et al.[3] identified 767 protein family domains that are significantly enriched across multiple plant-associated taxa. Many of the genes they identified are predicted to be involved in carbon or carbohydrate sensing or utilization. These include LacI family transcription factors that regulate transcription in response to carbon sensing and an aldo-keto reductase family involved in breakdown of a variety of substrates including sugars and reactive carbonyl compounds. Levy et al.[3] also found a number of genes that appear eukaryotic-like, which they named PREPARADOs (plant-resembling plant-associated and root-associated domains), and it is tempting to speculate that these gene products interfere with or mimic host cell signaling. While the majority of the candidate genes identified in this study remain unvalidated, this is the largest such resource available thus far and is a tool for hypothesis generation about genes that drive adaptation to the plant rhizosphere.

From genomes to mechanisms

Levy et al.[3] performed preliminary characterization of several candidates predicted to be enriched in plant-associated taxa. Two gene families were enriched in plant-associated isolates of the Burkholderiales genus *Acidovorax*, which includes both commensals and pathogens. Within plant-associated *Acidovorax* isolates, these two gene families were strongly anticorrelated: pathogenic isolates of *Acidovorax* contained one gene family (named 'Hyde' by the authors) while non-pathogenic strains contained a second family ('Jekyll'). In the genomes of pathogenic *Acidovorax* strains, *Hyde* loci are located near type 6 secretion system structural genes known to be involved in cell–cell killing[8], and Levy et al.[3] propose that *Hyde* genes may function as type 6 effectors. They demonstrated that the *Hyde*-containing strains possess antibacterial activity that is lost in a Δ*Hyde* mutant. The authors speculate that *Hyde* genes may help eliminate competition by pathogens, thereby promoting the virulence of strains that contain the *Hyde* genes. Collectively, this work identified genes that are enriched in plant-associated bacterial taxa and in some cases, like *Jekyll*/*Hyde*, genes that correlate with behaviors like commensalism and pathogenicity.

Identifying the drivers of microbial lifestyle transitions is critical for controlling emergent pathogens and for microbiome engineering for agricultural improvement. Reductionist model systems have provided in-depth understanding of the molecular and genetic mechanisms that drive plant–microbe associations[9,10]. However powerful, these systems are limited in that they do not take into account the complex communities of microbes that associate with hosts in natural settings and tend to focus on a limited set of models for which extensive tools exist. Microbiome and bacterial genome sequencing have generated a massive quantity of data in an attempt to bridge the gap between these reductionist systems and the microbial complexity that exists in nature; however, it can be challenging to extract testable mechanistic hypotheses from large datasets. Levy et al.[3] leveraged the vast quantity of available bacterial genomic data to generate testable hypotheses about bacterial genes that drive plant association. The authors provided several case studies on how to combine their hypotheses with genetics in tractable microbes to generate plausible functions for several gene families. Ultimately, combining analyses like this one with model systems and synthetic communities has the potential to identify the molecular mechanisms that drive bacterial evolutionary transitions in ecologically and agriculturally important contexts[11,12]. ❐

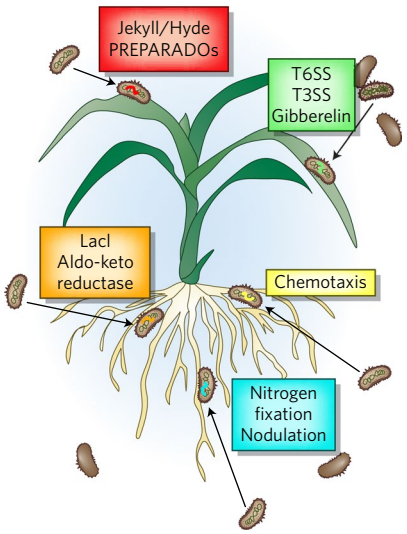

Fig. 1 | Identification of bacterial genes that are enriched in the genomes of plant-associated microbes. In a large-scale comparative genomics approach, Levy et al.[3] identified usual suspects involved in mutualistic or pathogenic associations with plants. This study also identified hundreds of novel gene families, including ones with potential roles in microbe–microbe competition and carbon utilization.

Ryan A. Melnyk[1] and Cara H. Haney[1,2]*
[1]*Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada.* [2]*Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada.*
**e-mail: cara.haney@msl.ubc.ca*

Published online: 22 December 2017
https://doi.org/10.1038/s41588-017-0019-2

References

1. Hacquard, S. et al. *Cell Host Microbe* **17**, 603–616 (2015).
2. Pel, M. J. C. & Pieterse, C. M. J. *J. Exp. Bot.* **64**, 1237–1248 (2013).
3. Levy, A. et al. *Nat. Genet.* https://doi.org/10.1038/s41588-017-0012-9 (2017).
4. Porter, S. S., Chang, P. L., Conow, C. A., Dunham, J. P. & Friesen, M. L. *ISME J.* **11**, 248–262 (2017).

5. Nowell, R. W., Laue, B. E., Sharp, P. M. & Green, S. *Mol. Plant Pathol.* **17**, 1409–1424 (2016).
6. Ali, S., Duan, J., Charles, T. C. & Glick, B. R. *J. Theor. Biol.* **343**, 193–198 (2014).
7. Loper, J. E. et al. *PLoS Genet.* **8**, e1002784 (2012).
8. Ho, B. T., Dong, T. G. & Mekalanos, J. J. *Cell Host Microbe* **15**, 9–21 (2014).
9. Xin, X.-F. & He, S. Y. *Annu. Rev. Phytopathol.* **51**, 473–498 (2013).
10. Kereszt, A., Mergaert, P., Maróti, G. & Kondorosi, E. *Curr. Opin. Microbiol.* **14**, 76–81 (2011).
11. Busby, P. E. et al. *PLoS Biol.* **15**, e2001793 (2017).
12. Vorholt, J. A., Vogel, C., Carlström, C. I. & Müller, D. B. *Cell Host Microbe* **22**, 142–155 (2017).

Competing interests
The authors declare no competing financial interests.

ENHANCER EPIGENETICS

Is H3K4me1 at enhancers correlative or causative?

H3K4me1 is enriched at active and primed enhancers. However, whether H3K4me1 controls or simply correlates with enhancer activity and function has remained unclear. Several recent reports, including two in *Nature Genetics*, provide major mechanistic and functional insights into the role of H3K4me1 at enhancers.

Alvaro Rada-Iglesias

In the last few years, epigenomic strategies have shown that genomes are filled with thousands of regulatory elements, such as enhancers, that are essential for the establishment of cell-type-specific gene expression programs[1]. Moreover, the general interest in enhancers has also increased owing to accumulating evidence suggesting that the genetic or structural disruption of enhancer function represents a major cause of human disease[2]. One powerful strategy to globally identify enhancers and to infer their regulatory state (for example, active, primed or poised) is to use distinctive chromatin signatures based on the enrichment profiles of several histone modifications[3–5]. Among these histone modifications, H3K4me1 is certainly important. First, coupled to H3K4me3 depletion, the presence of H3K4me1 distinguishes enhancers from proximal promoters[4]. Furthermore, H3K4me1 marks active and primed enhancers, which can be distinguished on the basis of the presence and absence of H3K27ac, respectively[3,5]. Consequently, a myriad of H3K4me1 profiles have been generated in different cell types and organisms to generate comprehensive enhancer maps. However, whether H3K4me1 has a regulatory function within enhancers or is simply a useful mark to identify them remained unclear (Fig. 1). Recent publications from the Ren, Shilatifard and Wysocka laboratories now provide major insights into the role of H3K4me1 at enhancers, suggesting overall that H3K4me1 might fine-tune, rather than tightly control, enhancer activity and function[6–8].

Enhancers and gene expression robustness

Histone modifications, such as H3K4me3, can serve as recruitment sites for protein complexes with major transcriptional regulatory functions[9]. On the basis of this idea, Local et al.[7] combined nucleosome pulldowns with mass spectrometry analysis to identify proteins that associate specifically with H3K4me1-containing nucleosomes. They observed specific interactions with members of several chromatin-remodeling complexes, which, at least in some cases (for example, the BAF complex), can control enhancer function[10]. Because most H3K4me1 at enhancers is deposited by KMT2C and KMT2D (MLL3 and MLL4, respectively)[11], Local et al.[7] used cells that were null or catalytically mutant for both these enzymes and that displayed highly reduced H3K4me1 levels. Importantly, loss of H3K4me1 resulted in decreased binding of BAF and other H3K4me1-associated proteins to enhancers. Moreover, using in vitro assays, they found that H3K4me1 increased the nucleosome-remodeling activity of the BAF complex. On the basis of these results, Local et al.[7] suggested that H3K4me1 is likely to have an active regulatory role at enhancers by serving as a docking site for chromatin remodelers.

Although the work from Local et al.[7] provides major mechanistic insights into the molecular function of H3K4me1 at enhancers, it did not directly test whether H3K4me1 is necessary for the activity of enhancers (for example, production of enhancer RNAs (eRNAs)) or for the expression of their target genes. Importantly, using mouse embryonic stem cells (mESCs) and *Drosophila melanogaster* as models, Dorigui et al.[6] and Rickels et al.[8] directly addressed these questions. Surprisingly, both studies found that, although H3K4me1 was severely reduced at enhancers in cells that were either null or catalytically mutant for KMT2C and KMT2D, only complete loss of both factors significantly affected enhancer activity and function. Therefore, they concluded that KMT2C and KMT2D, but not H3K4me1, have essential long-range regulatory functions in the context of active enhancers[6,8]. Is it possible to reconcile these results with the work from Local et al.[7]? KMT2C and KMT2D bind to and control the activity of a subset of but not all active enhancers[6–8]. Moreover, at this subset of enhancers, loss of H3K4me1 reduces but does not completely eliminate the recruitment of chromatin remodelers[7]. Considering that the expression of many genes seems to be controlled by multiple and potentially redundant enhancers[12], the previous observations suggest that loss of H3K4me1 might only partially reduce the total regulatory input that genes receive to sustain their expression. It is tempting to speculate that this partial reduction in enhancer regulatory input might preferentially affect gene expression under suboptimal conditions, such as those imposed by environmental perturbations[12]. Interestingly, Rickels et al.[8] found that KMT2C and KMT2D catalytically mutant flies were sensitive to environmental (high-temperature) and genetic (Nipped-B overexpression) perturbations, resulting in subtle phenotypes. Therefore, by facilitating the recruitment of chromatin remodelers, H3K4me1 might fine-tune enhancer activity,