Galaxy formation and evolution science in the era of the Large Synoptic Survey Telescope

Brant E. Robertson[1,2]*, *Manda Banerji*[3], *Sarah Brough*[4], *Roger L. Davies*[5], *Henry C. Ferguson*[6], *Ryan Hausen*[1], *Sugata Kaviraj*[7], *Jeffrey A. Newman*[8], *Samuel J. Schmidt*[9], *J. Anthony Tyson*[9] and *Risa H. Wechsler*[10]

Abstract | The field of galaxy formation and evolution synthesizes the physics of baryons and dark matter to describe the origin of systems such as the Milky Way and the enormous diversity of the galaxy population. The broad variation in possible formation histories and the wide range of cosmic environments make large statistical samples of galaxies essential for identifying the important physical mechanisms that govern their formation. Starting in the early 2020s, the Large Synoptic Survey Telescope (LSST) will provide an unmatched dataset for galaxy evolution studies by observing the entire southern sky in ultraviolet, optical and near-infrared wavelengths, producing multi-epoch digital images over a 10-year nominal mission that when summed will provide the deepest, wide-angle view of our Universe ever assembled. Here, we discuss the importance of LSST for deepening our understanding of galaxy formation and evolution over cosmic time. We present some outstanding problems in the field that LSST will address, and we present a roadmap of some preparatory research efforts required to make effective use of the LSST dataset for galaxy formation science.

Astronomy has moved into an era of large digital images of the sky from which enormous datasets including both the photometric properties of galaxies (photon flux in each of multiple bands, locations on the sky and redshift) and their structural characteristics (size and morphology) are measured. Compared with previous astronomical surveys, the Large Synoptic Survey Telescope (LSST) will provide a breakthrough in terms of the combined survey area and depth, the telescope mirror area and field of view (the étendue) and comparable advances in the quality and uniformity of the data processing and public access to the data[1]. A primary motivation for LSST has been to constrain cosmological parameters, including the energy density and possible time evolution of dark energy, through the weak gravitational lensing and angular clustering of galaxies[2]. This Roadmap discusses how LSST will also contribute to galaxy formation and evolution science, complementing the dark energy science that LSST will perform, and outlines some of the areas of active research that may influence how extragalactic astronomers can leverage the unprecedented dataset that the telescope will provide[3].

The features of LSST designed to help to optimize the survey for constraining dark energy will also enable its dataset to serve as a central driver for discoveries in galaxy formation physics. We provide an overview of the LSST facility and data relevant to understanding extragalactic science. Broadly, the important characteristics of LSST include a Wide–Fast–Deep (WFD) survey covering the entire southern sky in multiple bands from the ultraviolet to the near-infrared, with sufficient depth and imaging quality to reveal the detailed photometric and structural properties of billions of galaxies, and a complementary set of Deep Drilling Fields (DDFs) probing several times deeper than the main survey and covering areas ~10–100 times larger than any previous surveys at similar depth. LSST will probe the rarest environments where the highest matter overdensities or lowest underdensities may substantially affect the galaxy formation process. The image quality and depth will allow for a detailed understanding of how the process of galaxy formation establishes the full range of galaxy morphologies. The vast area of LSST will enable searches for the brightest and most massive galaxies and supermassive black holes (SMBHs) that cosmological structure formation can generate. LSST will thereby open up new lines of inquiry regarding galaxy formation, impossible to address with existing datasets, and may also lead to the resolution of long-standing puzzles in extragalactic science.

*e-mail: brant@ucsc.edu

https://doi.org/10.1038/s42254-019-0067-x

Key points

- The Large Synoptic Survey Telescope (LSST) will provide a new window into galaxy formation and evolution by imaging the entire southern sky with unprecedented sensitivity.
- By probing the rarest cosmic environments, LSST can reveal the extreme conditions under which the most luminous galaxies and the most massive supermassive black holes are formed.
- The gravitational lensing signals measured by LSST will enable astronomers to understand the mapping between observed galaxy properties and the dark matter halos that serve as the sites of galaxy formation.
- LSST will unveil the low-surface-brightness features around galaxies that encode the hierarchical nature of cosmological structure formation.
- A broad effort to address technical challenges for LSST, including deblending and machine learning, must commence now and requires heightened investment in the wide community of scientists who motivated the development of LSST.
- Coordinating existing ancillary data and new observational programmes in support of LSST will enable astronomers to make full use of the power of LSST, but these efforts need sufficient advance planning and community funding.

Deblending
The analysis procedure by which the light from overlapping stars and galaxies in crowded astronomical images is assigned to distinct objects.

Airmass
A measure of the path length through the Earth's atmosphere, increasing from zenith to the horizon.

Dithering strategies
The distribution on the sky of individual images taken by a camera, where shifts of the camera position are taken to cover more area, to improve the image quality by better sampling the light distribution of astronomical objects, and to account for the layout and possible defects of detectors in the camera.

Pointings
Single locations on the sky where a telescope has been aimed.

Although the LSST dataset will offer new opportunities, it also poses substantial challenges. The reduction of the data requires extreme care to preserve the faintest low-surface-brightness features that provide direct evidence of a hierarchical structure formation. The depth of the survey will lead to many objects of interest substantially overlapping with background objects, and the deblending of these superpositions requires careful attention to preserve the accuracy of the photometric measurements. The sheer volume and short cadence of the dataset present potential barriers to discovery unless astronomers can adapt existing tools and invent new methods to analyse efficiently trillions of data pixels. Advances in theoretical models and simulations must also keep pace with the improvements in the data, such that the full physical context of the observed populations can be understood. These challenges will lead to an exciting period of ingenuity and discovery over the next few years as astronomers prepare for the LSST science verification data in 2021 and full science operations starting in 2023.

Overview of the LSST

LSST represents a huge change in the combined data quality, area, wavelength coverage and depth available to astronomers (BOX 1). The LSST observatory uses a new, optimized telescope that combines a large-diameter mirror with a wide field of view and a fast optical design. The original optical design of the telescope was engineered roughly 20 years ago[4], with the measurement of the matter power spectrum through weak lensing as an original science driver[5]. The science portfolio for the observatory has expanded greatly since the LSST Science Book[2] and now reaches well beyond the stated focuses of dark matter and dark energy, near-Earth asteroids, transient astrophysics and the properties and structure of the Milky Way. While the important cosmological questions that LSST will explore[6] drive the scientific requirements (https://docushare.LSSTcorp.org/docushare/dsweb/Get/LPM-17) on the observatory[1] and analysis[7] side, extragalactic science exploring galaxy formation and evolution requires similar capabilities[3] and benefits substantially from the precision needed for the cosmological experiments.

To maximize the LSST survey area and depth given a nominal 10-year operations period, the efficient operation of the facility is required. To study how considerations including slew rates, exposure times, filter changes, weather patterns, moon phase and location, airmass and possible 'mini-surveys' with the telescope affect data acquisition and the overall survey efficiency, the LSST Project (https://www.LSST.org) developed an Operations Simulator[8] to model the execution of surveys with LSST. Based on quantifiable metrics that allow for the prioritization of science outcomes enabled by a given survey approach[9] and sky brightness models that set the exposure time requirements in each filter[10], the Operations Simulator has generated fiducial survey designs for LSST that demonstrate that the survey science requirements can be achieved and that science-driven optimizations of the survey design are possible. Indeed, the LSST community has contributed to the optimization of the LSST survey strategy, suggested dithering strategies to improve image quality and influenced the survey observing cadence through collaborative publications[11,12]. As the observatory is constructed, its expected performance is continuously being evaluated based on an integrated model of the telescope, optical path and facility[13].

Two important components of the LSST observing strategy include the WFDide–Fast–Deep main survey and set of deep individual pointings called the Deep Drilling Fields. The WFD survey will cover ~18,000 deg^2 in the *ugrizy* photometric bands to a depth of $r \approx 27.5$ AB magnitude[1]. Over the 10-year nominal survey period, each area on the sky will be visited 50–200 times in each band. By the end of the main WFD survey, LSST is expected to have catalogued ~37 billion objects across multiple epochs and identified $\gtrsim$7 trillion sources in one or more epochs (see LSST key numbers https://www.LSST.org/scientists/keynumbers). LSST will also observe at least four DDFs (known as ELAIS S1, XMM-LSS, ECDFS and COSMOS), comprising single ~9.6 deg^2 area pointings with *ugrizy* coverage to $r \approx 28$–28.5AB magnitude depth. LSST is currently considering other potential DDFs, mini-surveys or other adjustments to the survey strategy, but the balanced approach of the complementary WFD and DDFs will simultaneously achieve substantial depth over a considerable fraction of the whole sky and extreme depth over much larger areas than ever before acquired. The powerful capabilities of

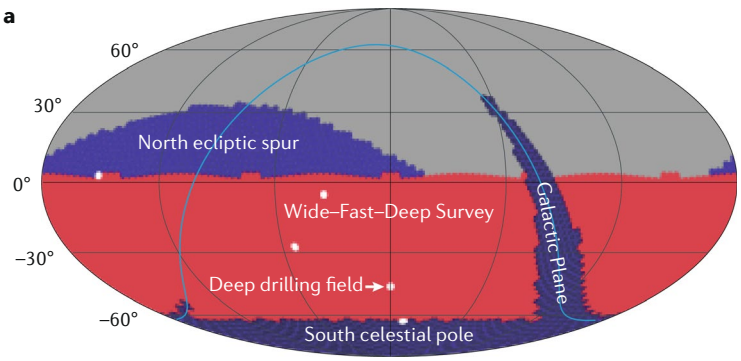
Author addresses

[1]Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA.
[2]Institute for Advanced Study, Princeton, NJ, USA.
[3]Institute of Astronomy, University of Cambridge, Cambridge, UK.
[4]School of Physics, University of New South Wales, Sydney, New South Wales, Australia.
[5]Department of Physics, University of Oxford, Oxford, UK.
[6]Space Telescope Science Institute, Baltimore, MD, USA.
[7]Centre for Astrophysics Research, University of Hertfordshire, Hatfield, UK.
[8]Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA, USA.
[9]Department of Physics, University of California, Davis, Davis, CA, USA.
[10]Department of Physics, Stanford University, Stanford, CA, USA.

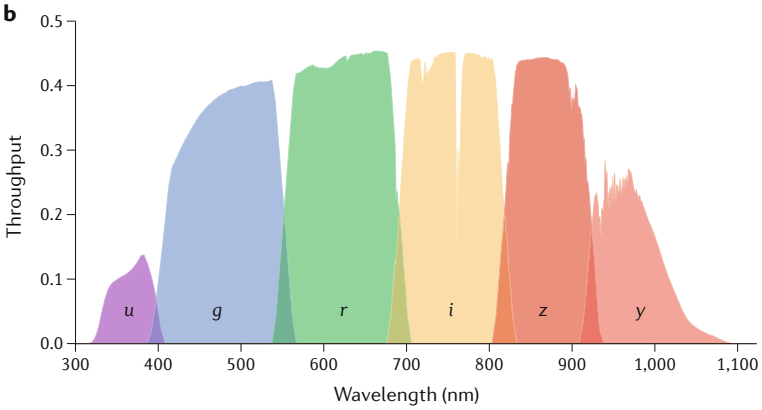
Box 1 | **The instrument**

The Large Synoptic Survey Telescope (LSST) will observe most of the southern sky over a 10-year period, producing an imaging dataset and source catalogue of great value for galaxy formation and evolution science. A Mollweide projection of the current baseline model for LSST (see the figure, panel **a**) shows example survey components. The primary Wide–Fast–Deep (WFD) Survey will reach a depth of $r \approx 27.5$ AB magnitude over ~18,000 deg^2 (the red area in panel **a**) and image in the *ugriz* filters (see the figure, panel **b**, for the total throughput, including atmosphere transmission and detector response with wavelength). The WFD Survey will detect more than 10 billion galaxies and enable position, flux, size and photometric redshift estimates for each source. This will be complemented by a collection of four or more Deep Drilling Fields (DDF; the white dot areas in panel **a**), each expected to cover a single LSST pointing (~9.6 deg^2) with *ugriz* imaging to $r \approx 28$–28.5AB magnitude depth. LSST will conduct shallower surveys of the south celestial pole and the Galactic Plane (the dark blue regions in panel **a**) that follows the midplane location of the Milky Way on the sky (the light blue line in panel **a**). LSST will also produce a shallow image of the north ecliptic spur region (the purple dot area in panel **a**), but most of the northern sky is inaccessible or difficult to observe from the LSST Cerro Pachón location in Chile and will not be imaged (the grey region in panel **a**). The LSST mirror size, field of view, camera size and the properties of the WFD and DDFs surveys are provided in the accompanying table in the figure, panel **c**.

a



b



c

Aperture size	8.4 m (6.4 m effective)
Field of view	9.6 deg^2
Camera	3.2 gigapixels
WFD Survey area	~18,000 deg^2
Expected WFD single-visit depths	$u = 23.9$, $g = 25.0$, $r = 24.7$, $i = 24.0$, $z = 23.3$, $y = 22.1$ (AB magnitudes)
Approximate WFD final depths	$u = 26.1$, $g = 27.4$, $r = 27.5$, $i = 26.8$, $z = 26.1$, $y = 24.9$ (AB magnitudes)
DDF Survey area	≥4 × 9.6 deg^2
Approximate DDF final depths	$r \approx 28$–28.5 AB magnitudes
Nominal operational lifetime	10 years

LSST are likely to be used for other surveys well after the end of the 10-year nominal operations period.

As LSST is a survey facility, the community has made substantial investments in the LSST Data Management System[14] that reduces the data and conducts basic photometric measurements. The data products produced by the LSST pipeline (see https://lse-163.LSST.io/) include Prompt products released daily to follow variable and Solar System objects, and Data Release products released yearly, including but not limited to calibrated single-epoch and co-added multi-epoch images and the photometric catalogue of objects. The catalogue entries will contain the object location, different measures of the multiband object flux including forward-modelled fluxes, sizes and accurate photometric redshift estimates[15] including representations of the corresponding redshift probability distribution function[16]. Users will access the data through the LSST Science Platform software framework, engineered as a set of online data analysis and visualization tools that allow astronomers to operate on the LSST imaging dataset and catalogues. These end-user analyses may lead to user-generated data products that extend the processing or augment the measurements provided by the LSST pipeline.

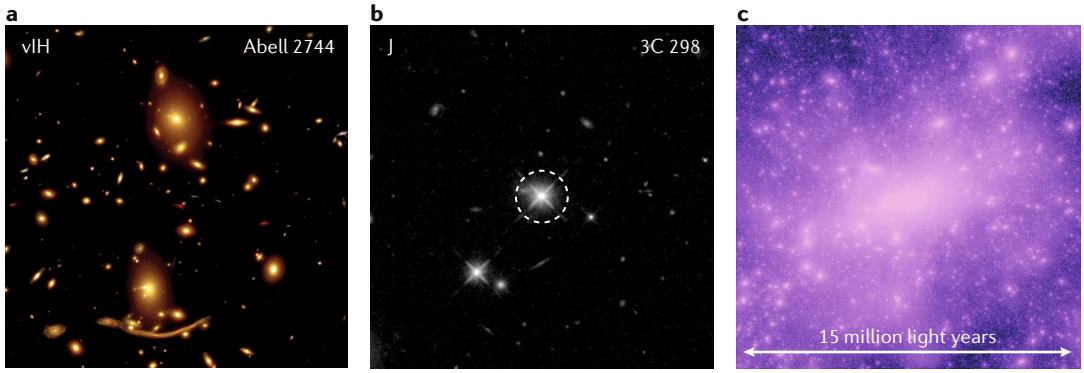
The community of scientists working to prepare for LSST observations includes hundreds of astronomers from all over the world. These scientists participate in LSST Science Collaborations in which astronomers interested in applications of LSST to domain science areas, such as galaxies (see LSST galaxies science collaboration https://galaxies.science.LSST.org/), organize preparatory research efforts. Members of the broader astronomical community[17], including at the national level in the United States, have considered what ancillary datasets and observing capabilities will lead to the largest science return from LSST[18] and are pursuing efforts to best leverage the investment in LSST by the United States government and other organizations.

Outstanding problems in galaxy formation

Galaxy formation and evolution involve physics across a vast range of scales[19] (BOX 2), but the bulk of the galaxy population follows regular trends in dynamical, structural, star formation, metallicities and other properties[20–32]. These trends primarily reflect a sequence of galaxy properties with mass but are further influenced by each object's individual star formation, merger history and environment[33–47]. The physics controlling these multivariate correlations involves a wide range of phenomena that serve energetically important roles on different scales, and probably include supernova explosions, radiation from young stars, energy input from accretion onto SMBHs, and processes related to hierarchical structure formation and gravitational clustering. Given that this complexity results in remarkably regular sequences with galaxy mass, the outcome of the galaxy formation process is often described by the connection between galaxy luminosity, or stellar mass, and their host dark matter halos[41,48–51]. Known as the stellar mass–halo mass relation, this relation succinctly captures how the physics of galaxy formation leads to an efficiency of star formation that varies with the host dark matter halo mass[52–55].

Box 2 | **Themes in galaxy formation and evolution**

The wide area, extensive multiband wavelength coverage and deep, high-quality imaging that the Large Synoptic Survey Telescope (LSST) will provide will enable astronomers to address important outstanding issues in galaxy formation and evolution. By surveying the entire southern sky, LSST will probe a wide range of cosmic structures and capture the rarest environments that affect the way galaxies form. Massive clusters, such as Abell 2744 shown in the figure, panel **a** in an image in the vIH filters from the Frontier Fields survey[64], influence their galaxy constituents through their hot gas that can shut down star formation in the satellite systems. LSST will discover even rarer cosmic environments and enable an understanding of how relations between local cosmic density and galaxy properties develop over time. The enormous area and multiband imaging of the survey allows for the identification of the rarest galaxies and most massive supermassive black holes (SMBHs) in the observable Universe. Quasars such as 3C 298, imaged in the infrared J-band with Hubble Space Telescope[206] (in the figure, panel **b**, source in the circle), are powered by accretion onto SMBHs and have been discovered as early as 700 million years after the Big Bang[75]. Such rare objects challenge our theories of galaxy formation because they have been assembled in such a short time that their formation efficiency must be very high. The formation of most galaxies is quite inefficient, which gives rise to a mass-dependent variation, for example, in the stellar mass content of a galaxy as a function of its dark matter halo mass. This galaxy–halo connection[55] reflects how the dark matter halo and substructures, as shown above on a scale of ~15 million light years selected from a larger cosmological simulation of structure formation (in the figure, panel **c**), map onto the observed distribution of galaxy luminosities. LSST can measure this relation directly by using weak gravitational lensing to infer the dark matter masses of galaxies[72,76]. Collectively, these science themes in galaxy formation and evolution will be areas of intense research as the LSST begins operations in 2023.



Imaging surveys have used complementary approaches to study the connection between galaxies and their host dark matter halos. Deep, space-based surveys with the Hubble Space Telescope including the Ultra Deep Field[56–59], Cosmological Evolution Survey (COSMOS)[60,61], Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS)[62,63] and Frontier Fields[64] have enabled the measurement of the rest-frame ultraviolet and optical luminosity functions over a wide span of cosmic history. The evolving ultraviolet luminosity density reflects the cosmic history of star formation[65], providing a global average of the bulk efficiency of galaxy formation with time. The deep surveys have shown that the co-moving density of the star formation rate declines from its peak at redshift $z \approx 2$ (about 10 billion years ago) to the present day, and the evolution in the high-redshift star formation at $z \approx 6–10$ is consistent with the re-ionization history[66] inferred from cosmic microwave background observations[67,68]. The high-quality photometry made possible by the deep data allows for well-constrained modelling of galaxy spectral energy distributions. It enables the generation of a histogram of stellar masses to be inferred from the spectral energy distribution fitting, and the 'stellar mass function' to be created. The image quality also allows morphological classification of the galaxies[69] and measurements of how the structural properties of galaxies evolve over time[70,71]. Weak lensing signals in the high-quality images provide statistical estimates of the gravitational masses of galaxies, and by combining with stellar mass estimates from the starlight, a direct connection between the stellar mass and halo mass of galaxies can be made[72].

Although space-based surveys provide exquisite resolution and depth, the modest areas they cover (less than approximately 1 deg^2) limit the information they provide on the connection between galaxies and their environment, rare objects and weak statistical signals that all require access to large areas of the sky. Wide-area, ground-based surveys that cover thousands of square degrees therefore provide complementary information to the deep, but narrow, space-based surveys. The relation between galaxies and their dark matter halos can also be constrained by their spatial clustering, which can be reliably modelled using cosmological simulations. The Sloan Digital Sky Survey has enabled precise galaxy clustering measurements combining both photometric and spectroscopic datasets[73]. Careful sample selection from the extensive source catalogues generated by wide-area surveys can allow the identification of rare and physically extreme sources, such as SMBHs in the very early Universe[74,75]. For gravitational lensing, existing surveys have enabled statistical detections of weak lensing and the measurement of the relation between galaxies and dark matter halos[76–79], complementing measurements from space-based imaging conducted on fewer objects but with higher data quality.

LSST will probe significantly larger areas more deeply than ever before, and therefore combines some of the

ugrizy photometric bands
The camera filters used by LSST to capture certain wavelengths of light, from the ultraviolet (u), to the optical (g, r and i) to the near-infrared (z and y).

Ancillary dataset
A collection of observations that augment or support an astronomical experiment, but were not taken by the same observatory (for instance, radio or X-ray observations provide ancillary datasets for optical observations).

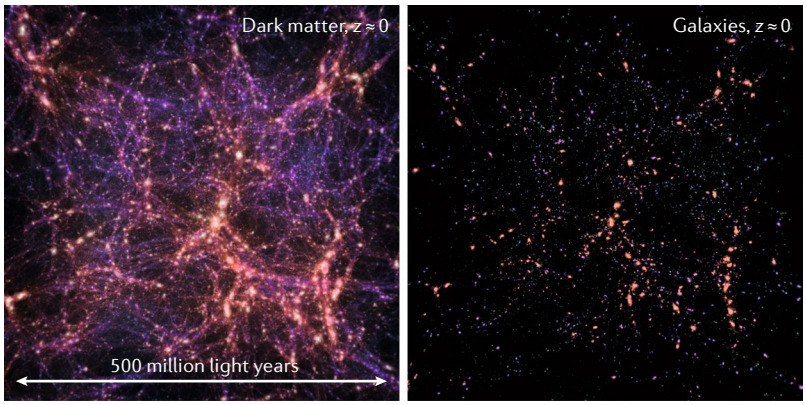


Fig. 1 | **Connection between dark matter structures and galaxies.** A prime aim for the Large Synoptic Survey Telescope (LSST) will be to measure the connection between dark matter structures and the galaxies that they host over a wide range of environments, scales and cosmic epochs. Theoretical models, such as the simulation shown in panel **a**, provide predictions for this correspondence that can be tested through observations. Panel **a** shows the results of a cosmological simulation for a number of particles $N = 2{,}048^3$ at the current epoch $z = 0$, modelling a cubic volume roughly 500 million light years on a side. The image intensity scales with the surface density of dark matter, and the coloration indicates the local dark matter velocity dispersion, increasing from dynamically cold regions (purple) to regions with large random motions reflecting deep gravitational potential wells (peach colour). The dark matter halos in this cosmological density field will host galaxies. This is illustrated in panel **b** in which galaxies are placed in regions dense enough to undergo gravitational collapse against the universal expansion and virialize. The colour of the galaxies reflects the mass surface density of their host dark matter halo, which would correlate with stellar mass through the stellar mass–halo mass relation. By measuring exquisite galaxy luminosity functions, spatial clustering and halo masses from weak gravitational lensing, LSST will reveal how the correspondence between galaxies and dark matter originates physically and whether it depends on cosmic environment.

best features of deep space-based surveys and wide-area ground-based surveys. The associated science gain is much greater than just the improved statistical constraints on galaxy populations, as LSST will enable an unmatched view into how the cosmic environment affects the formation of galaxies. Furthermore, the LSST data will contain the rarest, most extreme objects in the present-day Universe, and we will be able to use the unique depth and time baseline of LSST to learn about the most massive galaxies and SMBHs. The weak lensing signal from LSST, which is often considered only in the context of providing a constraint on dark energy, will reveal the detailed connection between galaxy properties and the masses of their dark matter halos. The abundance and structure of strong gravitational lenses in LSST will place new and powerful constraints on the nature of dark matter through the probes of halo substructure.

The wide and deep multiband photometry of LSST will enable fundamental new studies of the theoretical connection between galaxies and dark matter, of the low-surface-brightness signatures of hierarchical structure formation, and of the connection between SMBHs and galaxies. These topics are explored in more detail below. However, LSST cannot hope to explore all aspects of galaxy formation physics. The lack of a spectroscopic instrument means that the astrophysics associated with star formation and metal enrichment, supernova-driven winds, properties of the intergalactic medium and kinematic signatures of structure formation are unavailable to LSST on its own. To gain a fuller understanding of galaxy formation, LSST will need to connect with both spectroscopic and multiwavelength facilities that complement its capabilities. Some of these facilities and datasets are discussed below.

Theory of galaxy formation. The origin of the connection between galaxies and dark matter halos, and the role of the large-scale environment in setting the observable properties of galaxies[55] are among the most important unsolved problems in galaxy formation. Regarding the connection between galaxies and dark matter halos, surveys have not yet achieved definitive statistical estimates of weak lensing halo masses for the observed galaxy population[72,76,80–82]. Although the Dark Energy Survey and the Hyper Suprime-Cam survey are making substantial progress, the area and depth of LSST will greatly advance our understanding of how galaxies map to dark matter halos (FIG. 1) by probing the connection below the characteristic luminosity $L_\star$ of the galaxy luminosity function and extending previous results to higher redshifts[2]. Regarding the role of environment, the volumes probed by LSST and the extreme sensitivity of its combined dataset will place new constraints on our theories of galaxy formation and evolution. Where LSST is likely to prove most challenging for theoretical models is in the connection between galaxy formation, observed galaxy properties such as stellar mass and morphology, and the cosmic environment. As LSST will sample the most diverse range of environments ever observed, from the densest galaxy clusters to the rarest voids, it will not be easy to develop theories of galaxy formation that can explain regions with such widely varying densities and formation histories.

Theoretical modelling relevant for LSST includes many of the techniques currently used in cosmological simulations and semi-analytical models. Hydrodynamical simulations of galaxy formation continue to increase their volumes, improve their spatial resolution and include substantially more realistic physics than before[83–91]. Semi-analytical models have also improved their treatment of the physics of galaxy formation[92,93] and their computational speed[94], providing tools for understanding how the complexity of the galaxy formation process gives rise to the relatively simple observed dependence on the stellar mass. The results from these calculations can then funnel into sophisticated methods for simulating the observed data at the pixel level[95]. By comparing with the luminosity functions, stellar mass–halo mass relations and spatial clustering observed by LSST, theoretical models can tell us about the physics of galaxy formation. Luminosity functions constrain the rate at which galaxies form stars as a function of their stellar mass, which in turn constrains the energetics and coupling efficiency of feedback mechanisms associated with star formation (such as supernovae, radiation pressure and ionization). The stellar mass–halo mass relation constrains the overall efficiency of the galaxy formation process by measuring the fraction of the universal baryon content that forms

stars in structures with different dark matter masses. The spatial clustering of galaxies constrains the role of environment and the way in which luminous galaxies populate dark matter substructures within large dark matter halos. LSST will contribute to all these areas by providing the exacting statistical precision required to distinguish between stellar feedback models or models for the effects of environment on satellite galaxies, for instance.

The science requirements on simulations, such as those relevant for dark energy science with LSST[7], have pushed numerical techniques towards methods that can take advantage of the world's largest supercomputers. The need for precise values of dark matter halo mass functions[96], large-scale bias functions[97], volumes comparable to surveys of the entire sky, and many realizations to combat cosmic variance uncertainties[98] has motivated substantial cosmological efforts on leading supercomputers[99–101]. Such efforts have increasingly relied on exploiting new hardware architectures, such as graphics processor units, to realize the computational performance required to conduct the simulations in a reasonable time on even the most powerful computers[102]. Recently, these efforts have expanded to include hydrodynamic codes that run natively on the massively parallel architectures[103,104] that supply the vast majority of the computational brawn of modern supercomputers.

Further developmental efforts in simulation methodologies will need to connect the detailed physics modelled in high-resolution cosmological simulations with the volumes and statistical power of the simulations with the largest particle and cell numbers. This requirement holds especially true for understanding galaxy formation with LSST, as the simulations most useful for weak lensing studies typically lack the resolution or the hydrodynamics needed to model the detailed structure of individual galaxies. Current cosmological simulations with hydrodynamics do not cover a wide enough range of environments and peak rarity to characterize the diversity of galaxy formation pathways that LSST will probe. Resolving these competing interests will need the full computational power of new supercomputers, including the accelerated hardware they provide, while also combining the results of simulations with different resolutions, volumes and physical prescriptions to provide a hierarchy of calculations that are accurate and applicable on different scales.

Galaxies and supermassive black holes. SMBHs and galaxies share a fundamental link in their growth and evolution. The mass of SMBHs scales with the stellar mass and stellar velocity dispersion of their host galaxy spheroids[26,27,105–107], even as the physical scale of the galaxy extends well beyond the region in which the SMBH dominates the local gravitational potential. The galaxy–SMBH connection develops very early in cosmic history, as quasars, the most luminous active galactic nuclei (AGN), have been observed in the first billion years after the Big Bang[74,75,108–111]. SMBHs of up to ~10^9 solar masses power luminous quasars, and understanding how these enormous black holes grow in such a short time span remains a critical goal of galaxy formation theory[112]. Luminous early quasars probably grow in among the most massive early-forming galaxies[113]. As quasars are extremely rare on the sky, large-area surveys such as LSST are needed to find significant populations and cross-correlate them with the properties of nearby galaxies or cosmic environments.

LSST is expected to find more than 10^7 quasars[2], enabling a fundamentally new view into the SMBH–galaxy connection by allowing the construction of large quasar subsamples split by luminosity and environmental density, for instance. Photometric and synoptic optical/infrared surveys have two available methods for discovering quasars and AGN. Quasars discovered at progressively higher redshifts move through a multiband colour space, as lower-redshift Lyman-α absorption in the intergalactic medium and cosmic redshifting cause them to appear redder at higher redshift[108,114]. Optical AGN have long been known to show photometric variations[115–118], which can be modelled as a random walk and used to identify them in time-series data[119–121]. Both techniques will be used by LSST to identify quasars and AGN[2,122], measure their luminosity functions and spatial clustering, and probe the underlying connection between SMBHs and galaxies. The rapid cadence and nominal 10-year lifetime of LSST will enable AGN variability searches on timescales from a few days to years.

Variability can also be used to identify black holes in low-mass galaxies in the nearby Universe. By identifying AGN in low-mass galaxies, questions about the fraction of galaxies of any mass that host massive black holes and what the growth mechanisms of the smallest SMBHs that eventually seed higher luminosity AGN can be answered[123]. LSST will identify samples of low-luminosity AGN in small galaxies that can be followed up spectroscopically to determine how SMBHs continue to establish themselves on the SMBH–galaxy scaling relations, and why very low-mass galaxies and SMBHs may deviate from these relations[124–126].

Low-surface-brightness science. Low-surface-brightness features around galaxies encode information about how hierarchical structure formation builds galaxies through mergers and accretion. A primary quantity in galaxy formation and evolution involves the connection between the galaxy stellar mass and the mass of its host dark matter halo. The dark matter halo mass function is very steep, such that the abundance of halos increases rapidly at low masses[127,128]. Yet the abundance of low-mass galaxies does not follow the predicted halo abundance[129,130], and such galaxies contain significantly less than the universal fraction of baryons, reflecting a decrease of the stellar mass per unit dark matter halo mass. At the high-mass end of dark matter halos, in galaxy clusters in which a given halo may contain hundreds of galaxies populating individual sub-halo satellites, the ratio of stellar mass to halo mass again declines. The connection between dark matter halo and stellar mass in this regime encodes a variety of interesting galaxy formation physics, including the role of SMBHs, environment, gas accretion and dynamical interactions, but it has been difficult to constrain the stellar-to-halo

Peak rarity
A measure of the local overdensity of matter, equivalent to the number of standard deviations above the mean density a region would lie in the initial, nearly Gaussian matter density field generated by the end of inflation.

mass relation in this regime observationally. High-quality imaging over large survey areas is required, given the rarity of galaxy clusters, and, furthermore, the low-surface-brightness components of the brightest cluster galaxies must be seriously considered in determining the stellar mass assigned to very massive dark matter halos[51]. Without substantial care to understand intracluster light (ICL)[131] and the extended envelopes of galaxies with the largest stellar mass, the critical relation between stellar mass and halo mass can be misestimated. LSST will provide the definitive measurement of the stellar-to-halo mass relation in the high-mass limit, provided that its low-surface-brightness imaging capabilities can be maintained.

ICL is also interesting in its own right as it encodes the assembly history of galaxies in clusters, yet its properties, origin and formation over time are still an active area of research[131]. Clusters of galaxies are the most dynamically evolved, densest environments that galaxies occupy at every redshift. Interactions and coalescences will have been more frequent in clusters than anywhere else, and the material lost from individual galaxies will have been retained by the cluster. Thus the ICL arises from the stellar remnant of all the past interactions. The unprecedented depth achieved by LSST will allow us to measure the amount of light in the ICL (estimated to be 9–38% of the total[132]) and even determine the age and composition of the stars in the ICL. However, to do this, techniques must be developed to accurately estimate and remove the sky and Milky Way galactic cirrus background without removing the ICL as well.

Another crucial category of low-surface-brightness object involves the faint dwarfs that populate low-mass dark matter sub-halos in the Milky Way[133], and in galaxy groups and clusters. These objects were first discovered by using the Sloan Digital Sky Survey to search for stellar overdensities[134–136], and continue to be discovered using large-area photometric surveys such as the Dark Energy Survey, Kilo-Degree Survey and Hyper Suprime-Cam[137–140], and smaller facilities such as the Dragonfly Telephoto Array[141]. The completeness of searches for ultrafaint satellites with LSST will be influenced by its ability to optimize for the preservation of low-surface-brightness objects[2]. LSST is expected to discover a fair sample of these ultrafaint dwarfs[142] and may be able to use the corresponding galaxy counts to constrain the nature of dark matter by quantifying the matter power spectrum on scales below a billion solar masses.

Low-surface-brightness features encode the hierarchical character of galaxy formation, as individual accretion events in the stellar halo can be observed directly[143–146] and low-surface-brightness features around morphologically transformed galaxies indicate dynamical interactions[147,148]. Maintaining these features is not automatic in data pipelines, and requires substantial care in both sky subtraction and point spread function modelling[149–151], as many automated photometry programmes systematically over-subtract sky. LSST will probe to fainter surface-brightness limits over wider areas than ever before achievable. Maintaining proper sky subtraction to preserve the low-surface-brightness features around galaxies, such as stellar halos, tidal features and faint satellites, and to differentiate these from Milky Way galactic cirrus, will require tremendous care[152]. Considerable preparatory effort will be required to ensure that low-surface-brightness science is enabled while simultaneously accounting for instrumental features[153] such as stray light, ghosting and reflections from the telescope. LSST will reach ~30 mag arcsec^{-2} (magnitude per second of arc) in the 10-year stacked imaging, or possibly even deeper, but the final depth will depend sensitively on how the data are processed. As the depth and photometric quality of surveys have improved, the need to maintain the integrity of the low-surface-brightness data has become more critical.

Galactic cirrus background
A source of noise in the form of sky brightness from low density gas and dust in the Milky Way, spread over large angular scales on the sky.

Data analysis and technical challenges

LSST holds great promise for creating a definitive dataset for galaxy formation and evolution studies over the next decade and beyond. The realization of that promise will depend on a host of preparatory efforts to improve the analysis and calibration of the vast dataset of LSST images. Preparatory efforts need to extend into many areas of active research in galaxy formation and evolution, but the issues of addressing source deblending, applying machine learning to astronomical data, and optimizing synergy between LSST and other datasets deserve special attention in this Roadmap.

Deblending. The unique identification of objects in photometric surveys is critical for robustly measuring the luminosity of sources in each band, creating a baseline for the photometric variability that sources may exhibit, and inferring the structural properties of galaxies. LSST will provide unrivalled depth for a survey of its enormous area, but the sensitivity of the LSST imaging and the limitations of ground-based astronomy will result in the overlap of the observed surface-brightness distributions, or 'blending', of many sources in the data. Although the LSST location at the Cerro Pachón site in Chile shows excellent seeing (about 0.6″) and stability[154], LSST images will be seeing-limited, and not diffraction-limited as is possible over very small fields through the application of adaptive optics. Given the expected LSST seeing and the $r \approx 27.5$ AB depth of its WFD survey, more than half of all LSST sources are likely to show evidence for blending with neighbouring objects[155] (FIG. 2).

The blending of sources in the LSST data poses challenges for galaxy formation and evolution-related scientific analyses, and the topic is currently controversial. Any automated method for separating stars from galaxies must contend with overlapping objects. For blended galaxies, a robust method for separating the objects and assigning the observed flux in the overlapping pixels to individual objects must be engineered. Some objects cannot be easily identified as blends of multiple sources, depending on the photometric quality[156] of the ground-based data, and the inferred properties of such objects will be systematically in error. Even when sources are correctly identified as blends, the resulting photometric properties of the objects, including the flux, size and shape as measured by the data analysis pipeline, can be affected[157].

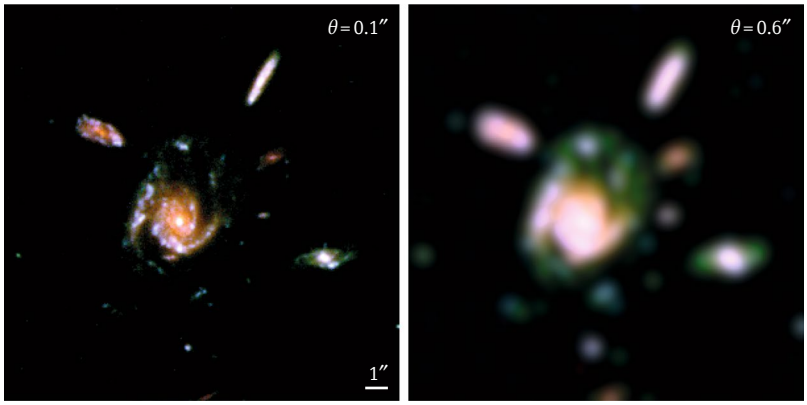


Fig. 2 | **Illustration of blending from atmospheric seeing.** The Large Synoptic Survey Telescope (LSST) will provide images of extraordinary depth for a wide-area survey, and in its DDFs it will reach $r \approx 28$–28.5, comparable to the GOODS-S Wide component of the Hubble Space Telescope survey Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS)[62,63]. Although LSST is a much larger telescope than Hubble, the ground-based LSST imaging will have a (still excellent) spatial resolution limited to $\theta \approx 0.6''$ by atmospheric turbulence. At the depths of the LSST surveys, this limitation will result in a substantial fraction of objects in LSST images being visual blends of multiple objects that could be distinctly identified with higher-resolution data[155]. Panel **a** shows a BVz colour image of a collection of galaxies in a reduction of the CANDELS GOODS-S image[59], showing the $\theta = 0.1''$ seeing of the Hubble Space Telescope unaffected by the atmosphere. Panel **b** shows the same region degraded by atmospheric turbulence to the $\theta \approx 0.6''$ seeing expected for LSST, which is excellent for a ground-based survey. In the seeing-limited data that LSST will acquire, the central spiral galaxy seen in panel **a** becomes blended with surrounding objects. Developing methods for the careful deblending of these LSST-quality images is an active area of preparatory research and is required to measure galaxy properties accurately in these very deep exposures[161].

Methodologies for source identification and deblending represent a very active research area in astronomical survey science[158–162]. Given the indeterminacy of how to assign flux uniquely from shared pixels to the multiple sources contributing to a blended object, forward modelling of surface-brightness distributions of sources subject to reasonable constraints has enabled some methods to achieve accurate model photometry for images with the degree of source blending expected for LSST. Substantial additional work would be required to enable these approaches in the LSST Data Management Pipeline[14], and catalogues using advances in deblending technology may need to be generated through post-processing efforts by the community. Preparatory research efforts may expand to make use of the colour information of objects, photometric redshifts, morphological classifications, space-based data as deblending priors or even machine learning methodologies to augment current deblending approaches. Although LSST cannot rival space-based facilities in the ability to deblend objects, and will therefore cede some science connected with galaxy morphology or low-mass satellite galaxies to those facilities, LSST data analysis methods must research deblending methodologies to make full use of its dataset.

Machine learning. With the advent of the LSST, the sheer size of the imaging dataset, expected to reach about 30 billion objects, will limit the ability to characterize qualitatively the properties of astronomical sources. The LSST pipeline will identify and measure the photometry for objects, providing a 'low-level' description of pixel-based quantities such as total flux or size, and 'intermediate-level' modelled quantities such as the best-fit Sersic[163] surface-brightness profile or photometric redshift. However, 'high-level' qualitative descriptions or classifications of objects that require a re-analysis of the imaging data after the LSST pipeline has completed will prove computationally challenging. Analysis tasks for which substantial advances will be necessary include the morphological classification of galaxies, the identification of tidal features and the association of interacting objects.

An important successful approach to overcoming galaxy morphological classification challenges posed by large imaging datasets has been the involvement and training of the public in citizen science projects such as Galaxy Zoo[164–166]. These efforts typically provide an online portal to train participants to identify and classify galaxies, for instance. The users are calibrated based on their accuracy and precision in their training, and then they are asked to perform classifications as they are able and willing. A critical aspect of this method relies on each object being classified by many people so that the range of classifications can be used to assign a confidence to the morphology of each galaxy. This approach has enabled the classification of morphology and other qualitative galaxy properties[167] for over a million objects from the Sloan Digital Sky Survey[168] and other datasets[169], but classifying objects at the scale of the entire LSST catalogue is likely to require other methods.

One class of technological solutions that has found success in astronomical classification problems is machine learning[170,171], and specifically its deep learning category involving convolutional neural networks[172]. Deep learning approaches have reached a classification accuracy on test datasets comparable to that of humans[171,173,174] (FIG. 3) and have been trained to perform other tasks in photometric pipelines such as surface-brightness profile fitting[175]. These methods have now been shown to transfer between different surveys[176], meaning that external precursor datasets may used to train deep learning models eventually applied to LSST data.

Because this field is still developing, there is substantial preparatory work ahead to make full use of machine learning methods for LSST. Machine learning methods remain controversial in the field, as they rely on complex computational methodologies that are not yet widely familiar to the astronomy community. Supervised methods require substantial training datasets that have been accurately classified, which most naturally requires collaboration between human and machine classifiers[177]. The most likely sources for the human classifications for training are citizen science efforts such as Galaxy Zoo or efforts involving the classification of more restricted datasets by professional astronomers[69]. The promise and efficiency of deep learning and other machine learning techniques lend themselves to other applications, such as photometric redshift determinations[178–180] or more even basic steps in the photometric pipeline such as object identification, deblending and segmentation.

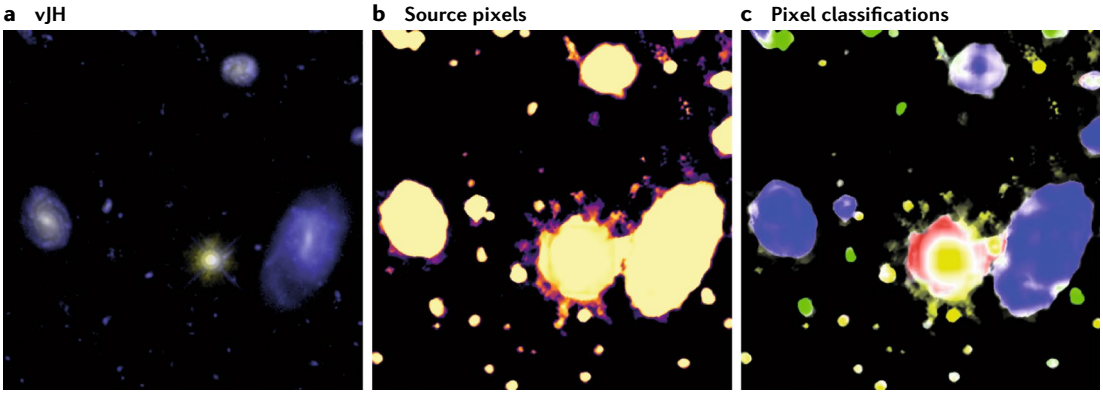


Fig. 3 | **Example of deep learning classification of astronomical image pixels.** Astronomical surveys with high-quality data can produce images that display the detailed morphology of galaxies. Panel **a** shows a vJH image of the GOODS-S region from the Hubble Space Telescope Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) programme[62,63]. The image in panel **a** contains a triad of blue galaxies that are clearly disks, a central yellow star and a variety of fainter objects with irregular or indeterminate morphologies. Using classifications by expert astronomers[69], a semantic segmentation model based on convolutional neural networks (R. Hausen, personal communication) was trained to reproduce simultaneously the expert astronomer classifications of galaxies and to identify which pixels in the image are associated with either sources or the blank sky. Panel **b** shows the model result for which pixels are likely to contain flux from sources in the image (brighter pixels have higher confidence). Panel **c** shows the per pixel classifications generated by the model, in which the blue pixels are associated with disk galaxies, green with irregular galaxies, yellow with point sources or unresolved objects, and red with spheroidal galaxies. Desaturated pixels indicate regions where the model has a high uncertainty about the pixel classification. The model accurately recovers the morphologies apparent in panel **a**, and the flux-weighted classification accuracy is excellent. Because these models process images quickly once trained, they can be deployed on large imaging datasets such as LSST to provide automated morphological classifications for billions of galaxies. Preparing these models for application to very large survey datasets is an area of active research in the field.

Precursor and ancillary datasets. In the decade leading up to LSST operations, a host of ground-based photometric and spectroscopic surveys have laid the ground work for the planning, analysis and discovery phases of LSST. For galaxy evolution studies, emphasis has been placed on obtaining multiwavelength datasets[181] in high-quality fields including the LSST DDFs[182]. There have been notable wide-area weak lensing surveys including the Dark Energy Survey[183,184], the Kilo-Degree Survey[185] and the Hyper Suprime-Cam Survey[186,187]. The VISTA (Visible and Infrared Survey Telescope for Astronomy) Hemisphere Survey[188] obtained Js-band and Ks-band near-infrared imaging that covers the southern sky. These precursor surveys approximate LSST in either area or depth and have many of the same scientific goals of LSST, including probing the galaxy–halo connection, AGN and galaxy clustering.

Wide-area spectroscopic surveys directly complement LSST by probing the astrophysics of ionized gas, metal line emission, properties of the intergalactic medium and galaxy kinematics that photometric surveys cannot directly observe. The Sloan Digital Sky Survey extended Baryon Oscillation Spectroscopic Survey[189], the Subaru Prime Focus Spectrograph Survey[190], the 4MOST (4-metre Multi-Object Spectroscopic Telescope) survey facility[191] and the Dark Energy Spectroscopic Instrument[192,193] will collectively supply spectra of tens of millions of galaxies over thousands of square degrees with partial overlap with the LSST footprint, and have conducted substantial but shallower imaging for target selection over areas comparable to LSST[194].

Whereas LSST can estimate galaxy stellar masses and star formation rates through photometric template fitting, observations of gas itself provide unique constraints on the galaxy baryon fraction, fuel for star formation and potentially rotational velocity measurements. In the radio, the Square Kilometer Array[195], and its pathfinders the Australian Square Kilometre Array Pathfinder[196] and MeerKAT[197], will provide complementary gas mass measurements that will augment the LSST rest-frame ultraviolet and optical imaging[198]. The eROSITA mission[199] will map the entire X-ray sky in 0.5–2 keV and 2–10 keV bands, supplying critical ancillary data on AGN and galaxy clusters in the LSST dataset.

LSST will find substantial synergy with two contemporary space-based missions[200]. The Euclid satellite mission[201], led by the European Space Agency, will provide another important space-based survey partially overlapping LSST[202] and will launch in 2021. Using a 1.2 m telescope with a ~0.54 deg^2 field-of-view, Euclid will cover ~15,000 deg^2 of sky to ~24.5AB magnitude depth in a broad visual filter and ~24AB magnitude depth in three near-infrared filters. The Euclid imaging survey will thereby enable weak lensing measurements of cosmic shear to constrain dark energy. The observatory will also perform slitless near-infrared spectroscopy to obtain redshifts for ~5×10^7 galaxies and achieve a high-accuracy measurement of baryonic acoustic oscillations. When combined with a pair of ~20 deg^2 deep pointings to ~26–26.5 AB magnitude depth, the Euclid survey will offer a rich dataset with superior resolution to the LSST seeing-limited imaging over a comparable area. The LSST data will probe substantially deeper than Euclid, and will supply multiband optical photometry that will greatly enhance the information content for galaxies observed with Euclid's single, wide R + I + Z visual filter.

Slitless near-infrared spectroscopy
A method for measuring the spectra of light from astronomical objects in a telescope's field of view by dispersing incoming light to a camera without using a mask with small slits.

R + I + Z visual filter
The camera filters used by Euclid to capture certain wavelengths of optical light.

H4RG infrared detectors
Special pixelated sensors used in astronomical cameras that have high quantum efficiency for light redder than wavelengths of 1 μm.

Similarly, in the mid-2020s NASA will launch the Wide Field InfraRed Survey Telescope (WFIRST), a satellite observatory with a 2.4 m primary mirror and an ~0.28 deg² field-of-view[203]. WFIRST will contain a bevy of H4RG infrared detectors that will enable ~26.7 AB magnitude ~1–2 μm imaging with supporting slitless spectroscopy over a cosmological survey covering ~2,000 deg² within the LSST WFD survey area, and will provide dark energy measurements through weak lensing and baryonic acoustic oscillations. WFIRST will also image ~20 deg² to ~28–29 AB magnitude depth through a supernova survey, well-matched and possibly co-located to an LSST DDF. The cosmological and supernova surveys will leverage the LSST optical imaging to provide accurate photometric redshifts for tens of millions of faint galaxies. Combined, LSST and WFIRST will provide a substantial archival resource for the astronomical community, with sensitive space-based infrared imaging to complement the LSST optical imaging in deep to ultra-deep areas spanning tens to thousands of square degrees. WFIRST will have a vibrant Guest Observer programme, enabling astronomers to follow up massive galaxies, rare quasars and ultra-rich galaxy clusters discovered in LSST data with high-quality space-based infrared imaging and spectroscopy[113].

Outlook

By providing a definitive multiband astronomical dataset that will shape ground-based astronomy for the foreseeable future, LSST will prove a transformative force in science. In addition to addressing an outstanding problem in physics — the nature and possible evolution of dark energy — and keeping a watchful eye on the threat from near-Earth asteroids[204], the scientific impact of LSST will extend throughout astrophysics. For the field of galaxy formation and evolution, the area of 18,000 deg² and depth of ~27.5 AB magnitude of the LSST WFD survey will enable an understanding of the connection between galaxy properties and cosmic environment, how structure formation in the cold dark matter paradigm allows for the most massive early galaxies and SMBHs to grow during the first billion years of cosmic history and the true diversity of formation pathways for galaxies like our Milky Way by the present day.

When combined with high-resolution imaging from space-based facilities such as WFIRST and Euclid, LSST will reveal how galaxy morphology developed over the past ~12 billion years and connect it with star formation rate, dark matter halo mass and environment. Wide-area infrared surveys will increase the precision of photometric redshifts from LSST and can substantially augment the LSST discovery space by improving constraints on stellar mass inferences from spectral energy distribution fitting[205], high-redshift galaxy identification and photometric redshift estimates. The LSST DDFs, comprising several areas of ~10 deg² located in regions with concentrated investments of ancillary data, will capture how the faint end of the galaxy luminosity distribution evolves over cosmic time, redistributing mass and luminosity increasingly towards brighter objects through physical mechanisms not yet fully known. As a result of the LSST data, galaxy formation and evolution as a field of astronomy will undergo an indelible change, as long-standing puzzles fall and we move closer to a more complete understanding of galaxy evolution.

A critical aspect to the change that LSST brings to science involves the ways through which astronomers conduct research. The LSST Project will release their entire dataset eventually, such that astronomers of all backgrounds from all over the world, professional or amateur, will enjoy access to the same premier dataset. With this breadth of access will come innovation, as different institutions and individual scientists will address galaxy formation and evolution with LSST through an ever wider range of approaches and goals. These studies will almost exclusively involve archival research efforts on the LSST WFD, DDFs or other pre-planned surveys, and their innovations will originate from creativity in how astronomers conduct their analysis rather than how they plan and execute their observations. Furthermore, because the acquisition of the LSST data will involve a dedicated and monumental effort over a decade, it will be a unique dataset. Correspondingly, reduction techniques, calibration, source identification, deblending and machine learning algorithms represent critical development areas that require heightened investment and effort starting now. These preparatory efforts that can make full use of the data that LSST acquires will attract increased research attention in coming years. As astronomers learn to cultivate the LSST data, discoveries hidden in the data will finally come to light.

Published online: 07 June 2019

1. Ivezić, Ž. et al. LSST: from science drivers to reference design and anticipated data products. *Astrophys. J.* **829**, 111 (2019).
Overview of the design specifications and the science requirements of LSST.
2. LSST Science Collaboration et al. *LSST Science Book*, version 2.0. Preprint at *arXiv* https://arxiv.org/abs/0912.0201 (2009).
Extensive collection of community-driven science cases motivating the construction of LSST.
3. Robertson, B. E. et al. Large synoptic survey telescope galaxies science roadmap. Preprint at *arXiv* https://arxiv.org/abs/1708.01617 (2017).
Detailed list of preparatory research activities and deliverables for the LSST Galaxies Science Collaboration.
4. Angel, R., Lesser, M., Sarlot, R. & Dunham, E. in *Imaging the Universe in Three Dimensions* (eds. van Breugel, W. & Bland-Hawthorn, J.) 81 (Conference Series no. 195, Astronomical Society of the Pacific, 2000).
5. Tyson, A. & Angel, R. in *The New Era of Wide Field Astronomy* (eds Clowes, R., Adamson, A. & Bromage, G.) 347 (Conference Series no. 232, Astronomical Society of the Pacific, 2001).
6. Zhan, H. & Tyson, J. A. Cosmology with the Large Synoptic Survey Telescope: an overview. *Rep. Prog. Phys.* **81**, 066901 (2018).
7. The LSST Dark Energy Science Collaboration et al. The LSST Dark Energy Science Collaboration (DESC) science requirements document. Preprint at *arXiv* https://arxiv.org/abs/1809.01669 (2018).
Detailed list of preparatory research activities and deliverables for the LSST Dark Energy Science Collaboration.
8. Reuter, M. A., Cook, K. H., Delgado, F., Petry, C. E. & Ridgway, S. T. Simulating the LSST OCS for conducting survey simulations using the LSST scheduler. In *Proc. SPIE 9911: Modeling, Systems Engineering, and Project Management for Astronomy VI*, https://doi.org/10.1117/12.2232680 (SPIE, 2016).
9. Jones, R. L. et al. The LSST metrics analysis framework (MAF). In *Proc. SPIE 9149: Observatory Operations: Strategies, Processes, and Systems V*, 91490B, https://doi.org/10.1117/12.2056835 (SPIE, 2014).
10. Yoachim, P. et al. An optical to IR sky brightness model for the LSST. In *Proc. SPIE 9910: Observatory Operations: Strategies, Processes, and Systems VI*, 99101A, https://doi.org/10.1117/12.2232947 (SPIE, 2016).
11. Awan, H. et al. Testing LSST dither strategies for survey uniformity and large-scale structure systematics. *Astrophys. J.* **829**, 50 (2016).
12. LSST Science Collaboration et al. Science-driven optimization of the LSST observing strategy. Preprint at *arXiv* https://arxiv.org/abs/1708.04058 (2017).
13. Xin, B. et al. Monitoring LSST system performance during construction. In *SPIE 10705: Modeling, Systems Engineering, and Project Management for Astronomy VIII*, 107050P, https://doi.org/10.1117/12.2313880 (SPIE, 2018).

14. Jurić, M. et al. The LSST data management system. In *Astronomical Data Analysis Software and Systems XXV* (eds Lorente, N. P. F., Shortridge, K. & Wayth, R.) 279 (Conference Series no. 512, Astronomical Society of the Pacific, 2017).
15. Graham, M. L. et al. Photometric redshifts with the LSST: evaluating survey observing strategies. *Astron. J.* **155**, 1 (2018).
16. Malz, A. I. et al. Approximating photo-*z* PDFs for large surveys. *Astron. J.* **156**, 35 (2018).
17. Brough, S., Akhlaghi, M., Bian, F., Glazebrook, K. & Kuehn, K. LSST and Australia. In *Astronomical Data Analysis Software and Systems XXV* (eds Lorente, N. P. F., Shortridge, K. & Wayth, R.) 667 (Conference Series no. 512, Astronomical Society of the Pacific, 2017).
18. Najita, J. et al. Maximizing science in the era of LSST: a community-based study of needed US capabilities. Preprint at *arXiv* https://arxiv.org/abs/1610.01661 (2016).
19. Somerville, R. S. & Davé, R. Physical models of galaxy formation in a cosmological framework. *Annu. Rev. Astron. Astrophys.* **53**, 51–113 (2015).
20. Faber, S. M. & Jackson, R. E. Velocity dispersions and mass-to-light ratios for elliptical galaxies. *Astrophys. J.* **204**, 668–683 (1976).
21. Tully, R. B. & Fisher, J. R. A new method of determining distances to galaxies. *Astron. Astrophys.* **54**, 661–673 (1977).
22. Djorgovski, S. & Davis, M. Fundamental properties of elliptical galaxies. *Astrophys. J.* **313**, 59–68 (1987).
23. Dressler, A. et al. Spectroscopy and photometry of elliptical galaxies. I: A new distance estimator. *Astrophys. J.* **313**, 42–58 (1987).
24. Roberts, M. S. & Haynes, M. P. Physical parameters along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **32**, 115–152 (1994).
25. Kennicutt, R. C. Jr. The global Schmidt law in star-forming galaxies. *Astrophys. J.* **498**, 541–552 (1998).
26. Gebhardt, K. et al. A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J. Lett.* **539**, L13–L16 (2000).
27. Ferrarese, L. & Merritt, D. A fundamental relation between supermassive black holes and their host galaxies. *Astrophys. J. Lett.* **539**, L9–L12 (2000).
28. Bell, E. F. & de Jong, R. S. Stellar mass-to-light ratios and the Tully–Fisher relation. *Astrophys. J.* **550**, 212–229 (2001).
29. Tremonti, C. A. et al. The origin of the mass–metallicity relation: insights from 53,000 star-forming galaxies in the Sloan Digital Sky Survey. *Astrophys. J.* **613**, 898–913 (2004).
30. Noeske, K. G. et al. Star formation in AEGIS field galaxies since $z = 1.1$: the dominance of gradually declining star formation, and the main sequence of star-forming galaxies. *Astrophys. J. Lett.* **660**, L43–L46 (2007).
31. Whitaker, K. E., van Dokkum, P. G., Brammer, G. & Franx, M. The star formation mass sequence out to $z = 2.5$. *Whitaker2012a* **754**, L29 (2012).
32. Salmon, B. et al. The relation between star formation rate and stellar mass for galaxies at $3.5 \leq z \leq 6.5$ in CANDELS. *Astrophys. J.* **799**, 183 (2015).
33. Cooper, M. C. et al. The DEEP2 galaxy redshift survey: the relationship between galaxy properties and environment at $z \sim 1$. *Mon. Not. R. Astron. Soc.* **370**, 198–212 (2006).
34. Daddi, E. et al. Multiwavelength study of massive galaxies at $z \sim 2$. I. Star formation and galaxy growth. *Astrophys. J.* **670**, 156–172 (2007).
35. Elbaz, D. et al. The reversal of the star formation–density relation in the distant Universe. *Astron. Astrophys.* **468**, 33–48 (2007).
36. Cooper, M. C. et al. The DEEP2 galaxy redshift survey: the role of galaxy environment in the cosmic star formation history. *Mon. Not. R. Astron. Soc.* **383**, 1058–1078 (2008).
37. Cooper, M. C., Tremonti, C. A., Newman, J. A. & Zabludoff, A. I. The role of environment in the mass–metallicity relation. *Mon. Not. R. Astron. Soc.* **390**, 245–256 (2008).
38. Peng, Y.-j et al. Mass and environment as drivers of galaxy evolution in SDSS and zCOSMOS and the origin of the Schechter function. *Astrophys. J.* **721**, 193–221 (2010).
39. Davé, R., Oppenheimer, B. D. & Finlator, K. Galaxy evolution in cosmological simulations with outflows — I. Stellar masses and star formation rates. *Mon. Not. R. Astron. Soc.* **415**, 11–31 (2011).
40. Davé, R., Finlator, K. & Oppenheimer, B. D. Galaxy evolution in cosmological simulations with outflows — II. Metallicities and gas fractions. *Mon. Not. R. Astron. Soc.* **416**, 1354–1376 (2011).
41. Behroozi, P. S., Wechsler, R. H. & Conroy, C. The average star formation histories of galaxies in dark matter halos from $z = 0$–8. *Astrophys. J.* **770**, 57 (2013).
42. Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A. & Peng, Y. Gas regulation of galaxies: the evolution of the cosmic specific star formation rate, the metallicity–mass–star-formation rate relation, and the stellar content of halos. *Astrophys. J.* **772**, 119 (2013).
43. Brough, S. et al. The SAMI galaxy survey: mass as the driver of the kinematic morphology–density relation in clusters. *Astrophys. J.* **844**, 59 (2017).
44. Martin, G. et al. The limited role of galaxy mergers in driving stellar mass growth over cosmic time. *Mon. Not. R. Astron. Soc.* **472**, L50–L54 (2017).
45. Weigel, A. K. et al. Galaxy zoo: major galaxy mergers are not a significant quenching pathway. *Astrophys. J.* **845**, 145 (2017).
46. Martin, G., Kaviraj, S., Devriendt, J. E. G., Dubois, Y. & Pichon, C. The role of mergers in driving morphological transformation over cosmic time. *Mon. Not. R. Astron. Soc.* **480**, 2266–2283 (2018).
47. Wang, L. et al. Galaxy and Mass Assembly (GAMA): the environmental dependence of the galaxy main sequence. Preprint at *arXiv* https://arxiv.org/abs/1802.08456 (2018).
48. Vale, A. & Ostriker, J. P. Linking halo mass to galaxy luminosity. *Mon. Not. R. Astron. Soc.* **353**, 189–200 (2004).
49. Kravtsov, A. V. et al. The dark side of the halo occupation distribution. *Astrophys. J.* **609**, 35–49 (2004).
50. Moster, B. P., Naab, T. & White, S. D. M. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *Mon. Not. R. Astron. Soc.* **428**, 3121–3138 (2013).
51. Kravtsov, A. V., Vikhlinin, A. A. & Meshcheryakov, A. V. Stellar mass–halo mass relation and star formation efficiency in high-mass halos. *Astron. Lett.* **44**, 8–34 (2018).
52. Conselice, C. J., Twite, J. W., Palamara, D. P. & Hartley, W. The halo masses of galaxies to $z \sim 3$: a hybrid observational and theoretical approach. *Astrophys. J.* **863**, 42 (2018).
53. Moster, B. P., Naab, T. & White, S. D. M. EMERGE — an empirical model for the formation of galaxies since $z \sim 10$. *Mon. Not. R. Astron. Soc.* **477**, 1822–1852 (2018).
54. Behroozi, P., Wechsler, R., Hearin, A. & Conroy, C. UniverseMachine: the correlation between galaxy growth and dark matter halo assembly from $z = 0$–10. Preprint at *arXiv* https://arxiv.org/abs/1806.07893 (2018).
55. Wechsler, R. H. & Tinker, J. L. The connection between galaxies and their dark matter halos. Preprint at *arXiv* https://arxiv.org/abs/1804.03097 (2018).
Review of the physics driving the relationship between galaxies, their observable properties and their host dark matter halos.
56. Beckwith, S. V. W. et al. The Hubble Ultra Deep Field. *Astron. J.* **132**, 1729–1755 (2006).
57. Ellis, R. S. et al. The abundance of star-forming galaxies in the redshift range 8.5–12: new results from the 2012 Hubble Ultra Deep Field campaign. *Astrophys. J. Lett.* **763**, L7 (2013).
58. Koekemoer, A. M. et al. The 2012 Hubble Ultra Deep Field (UDF12): observational overview. *Astrophys. J. Suppl.* **209**, 3 (2013).
59. Illingworth, G. D. et al. The HST eXtreme Deep Field (XDF): combining all ACS and WFC3/IR data on the HUDF region into the deepest field ever. *Astrophys. J. Suppl.* **209**, 6 (2013).
60. Scoville, N. et al. The Cosmic Evolution Survey (COSMOS): overview. *Astrophys. J. Suppl.* **172**, 1–8 (2007).
61. Laigle, C. et al. The COSMOS2015 catalog: exploring the $1 < z < 6$ Universe with half a million galaxies. *Astrophys. J. Suppl.* **224**, 24 (2016).
62. Grogin, N. A. et al. CANDELS: the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. *Astrophys. J. Suppl.* **197**, 35 (2011).
63. Koekemoer, A. M. et al. CANDELS: the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey — the Hubble Space Telescope observations, imaging data products, and mosaics. *Astrophys. J. Suppl.* **197**, 36 (2011).
64. Lotz, J. M. et al. The Frontier Fields: survey design and initial results. *Astrophys. J.* **837**, 97 (2017).
65. Madau, P. & Dickinson, M. Cosmic star-formation history. *Annu. Rev. Astron. Astrophys.* **52**, 415–486 (2014).
66. Robertson, B. E., Ellis, R. S., Furlanetto, S. R. & Dunlop, J. S. Cosmic reionization and early star-forming galaxies: a joint analysis of new constraints from Planck and the Hubble Space Telescope. *Astrophys. J. Lett.* **802**, L19 (2015).
67. Planck Collaboration et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594**, A13 (2016).
68. Planck Collaboration et al. Planck 2018 results. VI. Cosmological parameters. Preprint at *arXiv* https://arxiv.org/abs/1807.06209 (2018).
69. Kartaltepe, J. S. et al. CANDELS visual classifications: scheme, data release, and first results. *Astrophys. J. Suppl.* **221**, 11 (2015).
70. Shibuya, T., Ouchi, M. & Harikane, Y. Morphologies of ~190,000 galaxies at $z = 0$–10 revealed with HST legacy data. I. Size evolution. *Astrophys. J. Suppl.* **219**, 15 (2015).
71. Oldham, L. J., Houghton, R. C. W. & Davies, R. L. The most massive galaxies in clusters are already fully grown at $z \sim 0.5$. *Mon. Not. R. Astron. Soc.* **465**, 2101–2119 (2017).
72. Leauthaud, A. et al. New constraints on the evolution of the stellar-to-dark matter connection: a combined analysis of galaxy–galaxy lensing, clustering, and stellar mass functions from $z = 0.2$ to $z = 1$. *Astrophys. J.* **744**, 159 (2012).
73. Alam, S. et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. R. Astron. Soc.* **470**, 2617–2652 (2017).
74. Fan, X. et al. Constraining the evolution of the ionizing background and the epoch of reionization with $z \sim 6$ quasars. II. A sample of 19 quasars. *Astron. J.* **132**, 117–136 (2006).
75. Bañados, E. et al. An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. *Nature* **553**, 473–476 (2018).
76. Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M. & Brinkmann, J. Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies. *Mon. Not. R. Astron. Soc.* **368**, 715–731 (2006).
77. Heymans, C. et al. CFHTLenS: the Canada–France–Hawaii telescope lensing survey. *Mon. Not. R. Astron. Soc.* **427**, 146–166 (2012).
78. van Uitert, E. et al. The stellar-to-halo mass relation of GAMA galaxies from 100 deg^2 of KiDS weak lensing data. *Mon. Not. R. Astron. Soc.* **459**, 3251–3270 (2016).
79. Mandelbaum, R. et al. The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru strategic program survey. *Publ. Astron. Soc. Jpn.* **70**, S25 (2018).
80. Hoekstra, H., Hsieh, B. C., Yee, H. K. C., Lin, H. & Gladders, M. D. Virial masses and the baryon fraction in galaxies. *Astrophys. J.* **635**, 73–85 (2005).
81. Heymans, C. et al. A weak lensing estimate from GEMS of the virial to stellar mass ratio in massive galaxies to $z \sim 0.8$. *Mon. Not. R. Astron. Soc.* **371**, L60–L64 (2006).
82. Han, J. et al. Galaxy and Mass Assembly (GAMA): the halo mass of galaxy groups from maximum-likelihood weak lensing. *Mon. Not. R. Astron. Soc.* 446, (1356–1379 (2015).
83. Vogelsberger, M. et al. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature* **509**, 177–182 (2014).
84. Vogelsberger, M. et al. Introducing the Illustris project: simulating the coevolution of dark and visible matter in the Universe. *Mon. Not. R. Astron. Soc.* **444**, 1518–1547 (2014).
85. Dubois, Y. et al. Dancing in the dark: galactic properties trace spin swings along the cosmic web. *Mon. Not. R. Astron. Soc.* **444**, 1453–1468 (2014).
86. Schaye, J. et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. R. Astron. Soc.* **446**, 521–554 (2015).
87. Crain, R. A. et al. The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. *Mon. Not. R. Astron. Soc.* **450**, 1937–1961 (2015).
88. Feng, Y. et al. The BlueTides simulation: first galaxies and reionization. *Mon. Not. R. Astron. Soc.* **455**, 2778–2791 (2016).

89. Kaviraj, S. et al. The Horizon-AGN simulation: evolution of galaxy properties over cosmic time. *Mon. Not. R. Astron. Soc.* **467**, 4739–4752 (2017).
90. Di Matteo, T., Croft, R. A. C., Feng, Y., Waters, D. & Wilkins, S. The origin of the most massive black holes at high-*z*: BlueTides and the next quasar frontier. *Mon. Not. R. Astron. Soc.* **467**, 4243–4251 (2017).
91. Springel, V. et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Mon. Not. R. Astron. Soc.* **475**, 676–698 (2018).
92. Benson, A. J. GALACTICUS: a semi-analytic model of galaxy formation. *New Astron.* **17**, 175–197 (2012).
93. Lacey, C. G. et al. A unified multiwavelength model of galaxy formation. *Mon. Not. R. Astron. Soc.* **462**, 3854–3911 (2016).
94. Lagos, Cd. P. et al. Shark: introducing an open source, free, and flexible semi-analytic model of galaxy formation. *Mon. Not. R. Astron. Soc.* **481**, 3573–3603 (2018).
95. Rowe, B. T. P. et al. GALSIM: the modular galaxy image simulation toolkit. *Astron. Comput.* **10**, 121–150 (2015).
96. Tinker, J. et al. Toward a halo mass function for precision cosmology: the limits of universality. *Astrophys. J.* **688**, 709–728 (2008).
97. Tinker, J. L. et al. The large-scale bias of dark matter halos: numerical calibration and model tests. *Astrophys. J.* **724**, 878–886 (2010).
98. Garrison, L. H. et al. The Abacus cosmos: a suite of cosmological *N*-body simulations. *Astrophys. J. Suppl.* **236**, 43 (2018).
99. Heitmann, K., White, M., Wagner, C., Habib, S. & Higdon, D. The Coyote Universe. I. Precision determination of the nonlinear matter power spectrum. *Astrophys. J.* **715**, 104–121 (2010).
100. Heitmann, K., Lawrence, E., Kwan, J., Habib, S. & Higdon, D. The Coyote Universe extended: precision emulation of the matter power spectrum. *Astrophys. J.* **780**, 111 (2014).
101. Heitmann, K. et al. The Mira–Titan Universe: precision predictions for dark energy surveys. *Astrophys. J.* **820**, 108 (2016).
102. Habib, S. et al. HACC: simulating sky surveys on state-of-the-art supercomputing architectures. *New Astron.* **42**, 49–65 (2016).
103. Schneider, E. E. & Robertson, B. E. CHOLLA: a new massively parallel hydrodynamics code for astrophysical simulation. *Astrophys. J. Suppl.* **217**, 24 (2015).
104. Schneider, E. E. & Robertson, B. E. Hydrodynamical coupling of mass and momentum in multiphase galactic winds. *Astrophys. J.* **834**, 144 (2017).
105. Magorrian, J. et al. The demography of massive dark objects in galaxy centers. *Astron. J.* **115**, 2285–2305 (1998).
106. Tremaine, S. et al. The slope of the black hole mass versus velocity dispersion correlation. *Astrophys. J.* **574**, 740–753 (2002).
107. Häring, N. & Rix, H.-W. On the black hole mass–bulge mass relation. *Astrophys. J.* **604**, L89–L92 (2004).
108. Fan, X. et al. A survey of $z > 5.8$ quasars in the Sloan Digital Sky Survey. I. Discovery of three new quasars and the spatial density of luminous quasars at $z \sim 6$. *Astron. J.* **122**, 2833–2849 (2001).
109. Becker, R. H. et al. Evidence for reionization at $z \sim 6$: detection of a gunn-peterson trough in a $z = 6.28$ quasar. *Astron. J.* **122**, 2850–2857 (2001).
110. Fan, X. et al. A survey of $z > 5.7$ quasars in the Sloan Digital Sky Survey. II. Discovery of three additional quasars at $z > 6$. *Astron. J.* **125**, 1649–1659 (2003).
111. Mortlock, D. J. et al. A luminous quasar at a redshift of $z = 7.085$. *Nature* **474**, 616–619 (2011).
112. Efstathiou, G. & Rees, M. J. High-redshift quasars in the Cold Dark Matter cosmogony. *Mon. Not. R. Astron. Soc.* **230**, 5P–11P (1988).
113. Robertson, B., Li, Y., Cox, T. J., Hernquist, L. & Hopkins, P. F. Photometric properties of the most massive high-redshift galaxies. *Astrophys. J.* **667**, 60–78 (2007).
114. Richards, G. T. et al. Spectral energy distributions and multiwavelength selection of type 1 quasars. *Astrophys. J.* **166**, 470–497 (2006).
115. Matthews, T. A. & Sandage, A. R. Optical identification of 3C 48, 3C 196, and 3C 286 with stellar objects. *Astrophys. J.* **138**, 30 (1963).
116. Ulrich, M.-H., Maraschi, L. & Urry, C. M. Variability of active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **35**, 445–502 (1997).
117. Vanden Berk, D. E. et al. The ensemble photometric variability of 25,000 quasars in the Sloan Digital Sky Survey. *Astrophys. J.* **601**, 692–714 (2004).
118. Sesar, B. et al. Exploring the variable sky with the Sloan Digital Sky Survey. *Astron. J.* **134**, 2236–2251 (2007).
119. Kelly, B. C., Bechtold, J. & Siemiginowska, A. Are the variations in quasar optical flux driven by thermal fluctuations? *Astrophys. J.* **698**, 895–910 (2009).
120. MacLeod, C. L. et al. Modeling the time variability of SDSS stripe 82 quasars as a damped random walk. *Astrophys. J.* **721**, 1014–1033 (2010).
121. Kozłowski, S. et al. Quantifying quasar variability as part of a general approach to classifying continuously varying sources. *Astrophys. J.* **708**, 927–945 (2010).
122. MacLeod, C. L. et al. Quasar selection based on photometric variability. *Astrophys. J.* **728**, 26 (2011).
123. Baldassare, V. F., Geha, M. & Greene, J. Identifying AGNs in low-mass galaxies via long-term optical variability. *Astrophys. J.* **868**, 152 (2018).
124. Greene, J. E., Ho, L. C. & Barth, A. J. Black holes in pseudobulges and spheroidals: a change in the black hole–bulge scaling relations at low mass. *Astrophys. J.* **688**, 159–179 (2008).
125. Reines, A. E. & Volonteri, M. Relations between central black hole mass and total galaxy stellar mass in the local Universe. *Astrophys. J.* **813**, 82 (2015).
126. Baldassare, V. F., Reines, A. E., Gallo, E. & Greene, J. E. X-ray and ultraviolet properties of AGNs in nearby dwarf galaxies. *Astrophys. J.* **836**, 20 (2017).
127. Press, W. H. & Schechter, P. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *Astrophys. J.* **187**, 425–438 (1974).
128. Sheth, R. K., Mo, H. J. & Tormen, G. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Mon. Not. R. Astron. Soc.* **323**, 1–12 (2001).
129. Moore, B. et al. Dark matter substructure within galactic halos. *Astrophys. J. Lett.* **524**, L19–L22 (1999).
130. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where are the missing galactic satellites? *Astrophys. J.* **522**, 82–92 (1999).
131. Montes, M. & Trujillo, I. Intracluster light at the frontier — II. The Frontier Fields clusters. *Mon. Not. R. Astron. Soc.* **474**, 917–932 (2018).
132. Rudick, C. S., Mihos, J. C. & McBride, C. K. The quantity of intracluster light: comparing theoretical and observational measurement techniques using simulated clusters. *Astrophys. J.* **732**, 48 (2011).
133. McConnachie, A. W. The observed properties of dwarf galaxies in and around the Local Group. *Astron. J.* **144**, 4 (2012).
134. Willman, B. et al. A new Milky Way dwarf galaxy in Ursa Major. *Astrophys. J. Lett.* **626**, L85–L88 (2005).
135. Belokurov, V. et al. A faint new Milky Way satellite in Bootes. *Astrophys. J. Lett.* **647**, L111–L114 (2006).
136. Belokurov, V. et al. Cats and dogs, hair and a hero: a quintet of new Milky Way companions. *Astrophys. J.* **654**, 897–906 (2007).
137. Bechtol, K. et al. Eight new Milky Way companions discovered in first-year Dark Energy Survey data. *Astrophys. J.* **807**, 50 (2015).
138. Drlica-Wagner, A. et al. Eight ultra-faint galaxy candidates discovered in year two of the Dark Energy Survey. *Astrophys. J.* **813**, 109 (2015).
139. van der Burg, R. F. J. et al. The abundance of ultra-diffuse galaxies from groups to clusters. UDGs are relatively more common in more massive haloes. *Astron. Astrophys.* **607**, A79 (2017).
140. Greco, J. P. et al. Illuminating low surface brightness galaxies with the Hyper Suprime-Cam survey. *Astrophys. J.* **857**, 104 (2018).
141. Abraham, R. G. & van Dokkum, P. G. Ultra-low surface brightness imaging with the Dragonfly telephoto array. *Publ. Astron. Soc. Pac.* **126**, 55 (2014).
142. Tollerud, E. J., Bullock, J. S., Strigari, L. E. & Willman, B. Hundreds of Milky Way satellites? Luminosity bias in the satellite luminosity function. *Astrophys. J.* **688**, 277–289 (2008).
143. Belokurov, V. et al. The field of streams: Sagittarius and its siblings. *Astrophys. J. Lett.* **642**, L137–L140 (2006).
144. Bullock, J. S. & Johnston, K. V. Tracing galaxy formation with stellar halos. I. Methods. *Astrophys. J.* **635**, 931–949 (2005).
145. Johnston, K. V. et al. Tracing galaxy formation with stellar halos. II. Relating substructure in phase and abundance space to accretion histories. *Astrophys. J.* **689**, 936–957 (2008).
146. Bell, E. F. et al. The accretion origin of the Milky Way's stellar halo. *Astrophys. J.* **680**, 295–311 (2008).
147. Kaviraj, S. Peculiar early-type galaxies in the Sloan Digital Sky Survey stripe 82. *Mon. Not. R. Astron. Soc.* **406**, 382–394 (2010).
148. Kaviraj, S. The importance of minor-merger-driven star formation and black hole growth in disc galaxies. *Mon. Not. R. Astron. Soc.* **440**, 2944–2952 (2014).
149. Duc, P.-A. et al. The ATLAS^{3D} project — IX. The merger origin of a fast- and a slow-rotating early-type galaxy revealed with deep optical imaging: first results. *Mon. Not. R. Astron. Soc.* **417**, 863–881 (2011).
150. Trujillo, I. & Fliri, J. Beyond 31 mag arcsec^{-2}: the frontier of low surface brightness imaging with the largest optical telescopes. *Astrophys. J.* **823**, 123 (2016).
151. Borlaff, A. et al. The missing light of the Hubble Ultra Deep Field. Preprint at *arXiv* https://arxiv.org/abs/1810.00002 (2018).
152. Ji, I., Hasan, I., Schmidt, S. J. & Tyson, J. A. Estimating sky level. *Publ. Astron. Soc. Pacif.* **130**, 084504 (2018).
153. Bradshaw, A. K., Lage, C. & Tyson, J. A. Characterization of LSST CCDs using realistic images, before first light. Preprint at *arXiv* https://arxiv.org/abs/1808.00534 (2018).
154. Gressler, W. et al. LSST telescope and site status. In *Ground-based and Airborne Telescopes V. Proc. SPIE* **9145**, https://doi.org/10.1117/12.2056711 (SPIE, 2014).
155. Bosch, J. et al. The Hyper Suprime-Cam software pipeline. *Publ. Astron. Soc. Jpn* **70**, S5 (2018).
156. Ivezić, Ž. et al. SDSS data management and photometric quality assessment. *Astron. Nachr.* **325**, 583–589 (2004).
157. Dawson, W. A., Schneider, M. D., Tyson, J. A. & Jee, M. J. The ellipticity distribution of ambiguously blended objects. *Astrophys. J.* **816**, 11 (2016).
158. Merlin, E. et al. T-PHOT: a new code for PSF-matched, prior-based, multiwavelength extragalactic deconfusion photometry. *Astron. Astrophys.* **582**, A15 (2015).
159. Merlin, E. et al. T-PHOT version 2.0: improved algorithms for background subtraction, local convolution, kernel registration, and new options. *Astron. Astrophys.* **595**, A97 (2016).
160. Joseph, R., Courbin, F. & Starck, J.-L. Multi-band morpho-spectral component analysis deblending tool (MuSCADeT): deblending colourful objects. *Astron. Astrophys.* **589**, A2 (2016).
161. Melchior, P. et al. SCARLET: source separation in multi-band images by constrained matrix factorization. *Astron. Comput.* **24**, 129–142 (2018).
162. Robotham, A. S. G. et al. ProFound: source extraction and application to modern survey data. *Mon. Not. R. Astron. Soc.* **476**, 3137–3159 (2018).
163. Sersic, J. L. *Atlas de Galaxias Australes* (European Southern Observatory, 1968).
164. Lintott, C. J. et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **389**, 1179–1189 (2008).
165. Lintott, C. et al. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Mon. Not. R. Astron. Soc.* **410**, 166–178 (2011).
166. Willett, K. W. et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **435**, 2835–2860 (2013).
167. Bamford, S. P. et al. Galaxy Zoo: the dependence of morphology and colour on environment. *Mon. Not. R. Astron. Soc.* **393**, 1324–1352 (2009).
168. York, D. G. et al. The Sloan Digital Sky Survey: technical summary. *Astron. J.* **120**, 1579–1587 (2000).
169. Simmons, B. D. et al. Galaxy Zoo: quantitative visual morphological classifications for 48 000 galaxies from CANDELS. *Mon. Not. R. Astron. Soc.* **464**, 4420–4447 (2017).
170. Ball, N. M. et al. Galaxy types in the Sloan Digital Sky Survey using supervised artificial neural networks. *Mon. Not. R. Astron. Soc.* **348**, 1038–1046 (2004).
171. Hocking, A., Geach, J. E., Sun, Y. & Davey, N. An automatic taxonomy of galaxy morphology using unsupervised machine learning. *Mon. Not. R. Astron. Soc.* **473**, 1108–1129 (2018).
172. Krizhevsky, A., Sutskever, I. & Hinton, G. E. in *Advances in Neural Information Processing Systems* 25 (eds Pereira, F.et al.) 1097–1105 (Curran, 2012).
173. Dieleman, S., Willett, K. W. & Dambre, J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon. Not. R. Astron. Soc.* **450**, 1441–1459 (2015).
An early application of deep learning techniques to the automated morphological classification of galaxies.

174. Dai, J.-M. & Tong, J. Galaxy morphology classification with deep convolutional neural networks. Preprint at *arXiv* https://arxiv.org/abs/1807.10406 (2018).
175. Tuccillo, D. et al. Deep learning for galaxy surface brightness profile fitting. *Mon. Not. R. Astron. Soc.* **475**, 894–909 (2018).
176. Domnguez Sánchez, H. et al. Knowledge transfer of deep learning for galaxy morphology from one survey to another. Preprint at *arXiv* https://arxiv.org/abs/1807.00807 (2018).
177. Beck, M. R. et al. Integrating human and machine intelligence in galaxy morphology classification tasks. *Mon. Not. R. Astron. Soc.* **476**, 5516–5534 (2018).
178. Masters, D. et al. Mapping the galaxy color–redshift relation: optimal photometric redshift calibration strategies for cosmology surveys. *Astrophys. J.* **813**, 53 (2015).
179. Bilicki, M. et al. Photometric redshifts for the Kilo-Degree Survey. Machine-learning analysis with artificial neural networks. *Astron. Astrophys.* **616**, A69 (2018).
180. Pasquet, J., Bertin, E., Treyer, M., Arnouts, S. & Fouchez, D. Photometric redshifts from SDSS images using a convolutional neural network. *Astron. Astrophys.* **621**, A26 (2019).
181. Driver, S. P. et al. Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. *Monthly Notices of the Royal Astronomical Society* 413, 971–995 (2011).
182. Davies, L. J. M. et al. Deep Extragalactic Visible Legacy Survey (DEVILS): motivation, design, and target catalogue. *Mon. Not. R. Astron. Soc.* **480**, 768–799 (2018).
183. Dark Energy Survey Collaboration et al. The Dark Energy Survey: more than dark energy — an overview. *Mon. Not. R. Astron. Soc.* **460**, 1270–1299 (2016).
184. Abbott, T. M. C. et al. The Dark Energy Survey data release 1. Preprint at *arXiv* https://arxiv.org/abs/1801.03181 (2018).
185. Hildebrandt, H. et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Mon. Not. R. Astron. Soc.* **465**, 1454–1498 (2017).
186. Aihara, H. et al. The Hyper Suprime-Cam SSP survey: overview and survey design. *Publ. Astron. Soc. Jpn* **70**, S4 (2018).
187. Aihara, H. et al. First data release of the Hyper Suprime-Cam Subaru strategic program. *Publ. Astron. Soc. Jpn* **70**, S8 (2018).
188. McMahon, R. G. et al. First scientific results from the VISTA Hemisphere Survey (VHS). *Messenger* **154**, 35–37 (2013).
189. Abolfathi, B. et al. The fourteenth data release of the Sloan Digital Sky Survey: first spectroscopic data from the extended Baryon Oscillation Spectroscopic Survey and from the second phase of the Apache Point Observatory galactic evolution experiment. *Astrophys. J. Suppl.* **235**, 42 (2018).
190. Takada, M. et al. Extragalactic science, cosmology, and galactic archaeology with the Subaru Prime Focus Spectrograph. *Publ. Astron. Soc. Jpn* **66**, R1 (2014).
191. de Jong, R. S. et al. 4MOST: the 4-metre multi-object spectroscopic telescope project at preliminary design review. In *Proc. SPIE 9908: Ground-based and Airborne Instrumentation for Astronomy VI*, 99081 O (SPIE, 2016).
192. DESI Collaboration et al. The DESI experiment Part I: science, targeting, and survey design. Preprint at *arXiv* https://arxiv.org/abs/1611.00036 (2016).
193. DESI Collaboration et al. The DESI experiment Part II: instrument design. Preprint at *arXiv* https://arxiv.org/abs/1611.00037 (2016).
194. Dey, A. et al. Overview of the DESI Legacy Imaging Surveys. Preprint at *arXiv* https://arxiv.org/abs/1804.08657 (2018).
195. Carilli, C. L. & Rawlings, S. Motivation, key science projects, standards and assumptions. *New Astron. Rev.* **48**, 979–984 (2004).
196. Johnston, S. et al. Science with ASKAP. The Australian Square-Kilometre-Array Pathfinder. *Exp. Astron.* **22**, 151–273 (2008).
197. Jonas, J. L. MeerKAT — The South African array with composite dishes and wide-band single pixel feeds. *Proc. IEEE* **97**, 1522–1530 (2009).
198. Bacon, D. et al. Synergy between the Large Synoptic Survey telescope and the Square Kilometre Array. In *Proc. Advancing Astrophysics with the Square Kilometre Array. PoS(AASKA14)*, https://doi.org/10.22323/1.215.0145 (2015).
199. Merloni, A. et al. eROSITA science book: mapping the structure of the energetic Universe. Preprint at *arXiv* https://arxiv.org/abs/1209.3114 (2012).
200. Jain, B. et al. The whole is greater than the sum of the parts: optimizing the joint science return from LSST, Euclid and WFIRST. Preprint at *arXiv* https://arxiv.org/abs/1501.07897 (2015).
201. Laureijs, R. et al. Euclid definition study report. Preprint at *arXiv* https://arxiv.org/abs/1110.3193 (2011).
202. Rhodes, J. et al. Scientific synergy between LSST and Euclid. *Astrophys. J. Suppl.* **233**, 21 (2017).
203. Spergel, D. et al. Wide-field InfraRed survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 report. Preprint at *arXiv* https://arxiv.org/abs/1503.03757 (2015).
204. Jones, R. L. et al. The Large Synoptic Survey Telescope as a near-Earth object discovery machine. *Icarus* **303**, 181–202 (2018).
205. Banerji, M. et al. Combining Dark Energy Survey science verification data with near-infrared data from the ESO VISTA Hemisphere Survey. *Mon. Not. R. Astron. Soc.* **446**, 2523–2539 (2015).
206. Hilbert, B. et al. Powerful activity in the bright ages. I. A visible/IR survey of high redshift 3C radio galaxies and quasars. *Astrophys. J. Suppl.* **225**, 12 (2016).

Acknowledgements
B.E.R. acknowledges a Maureen and John Hendricks Visiting Professorship at the Institute for Advanced Study, NASA contract NNG16PJ25C, and NSF award 1828315. S.K. acknowledges support from STFC through grant ST/N002512/1 and a Senior Research Fellowship at Worcester College Oxford. S.J.S. acknowledges support from DOE grant DE-SC0009999 and NSF/AURA grant N56981C.

Author contributions
B.E.R. wrote the text and created the figures. M.B., S.B., R.L.D., H.C.F., R.H., S.K., J.A.N., S.J.S., J.A.T. and R.H.W. contributed to the text. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.