27 - PART 20 Emerging Topics in Clinical Medicine

01 - 493 Point-of-Care Ultrasound

493 Point-of-Care Ultrasound

Emerging Topics in Clinical Medicine PART 20 Wilma Chan, Nilam J. Soni, Paul H. Mayo

Point-of-Care Ultrasound DEFINITION Point-of-care ultrasound (POCUS) is defined as the acquisition, interpretation, and clinical integration of ultrasonographic views by a treating clinician in real time at the patient's bedside. POCUS is distinct from consultative ultrasound where a clinician orders an ultrasound exam, a sonographer acquires a comprehensive set of images, an imag ing specialist (most often a radiologist or cardiologist) interprets the images, and the ordering clinician receives an ultrasound report and integrates findings into clinical decision-making (Fig. 493-1). The goal of POCUS is not to replace the imaging specialist or the highresolution data provided by computed tomography (CT) or magnetic resonance imaging (MRI), but rather to improve diagnostic and thera peutic decisions made by the treating clinician at the bedside. POCUS became part of trauma care in emergency departments in the 1980s, and subsequently, many specialties began incorporat ing POCUS into patient care. The 1999 House of Delegates from the American Medical Association passed a resolution (AMA HR. 802) enabling each specialty to define its own scope and appropriate use of POCUS. Specialty-based guidelines emerged supporting credentialing processes and defining standard scanning protocols to answer focused diagnostic questions. Common clinical scenarios, such as acute dys pnea, abdominal pain, and shock, can be rapidly characterized using POCUS (Table 493-1). In internal medicine, there has been expanding interest in POCUS since the 2000s. POCUS can enhance diagnostic accuracy, treatment, monitoring, and screening of patients, as well as improve patient and clinician confidence and procedural safety (Fig. 493-2). Physical Examination Point-of-Care Consultative Ultrasound Ultrasound ask select acquire interpret act Bedside Clinician Sonographer Radiologist or Cardiologist FIGURE 493-1 Workflow schematic comparing point-of-care ultrasound to physical examination and consultative ultrasound. Medical decision-making begins with asking a targeted question, selecting the diagnostic modalities, acquiring and interpreting images or other data, and ultimately, acting to incorporate the new findings into the patient's care. The three different shapes represent various personnel in this process, and curved arrows demonstrate the exchange of information among providers across different stages. (Reproduced with permission from NJ Soni, BP Lucas: Diagnostic point-of-care ultrasound for hospitalists. J Hosp Med 10, 2014.)

Portable ultrasound machines are categorized as cart-based machines versus handheld devices with wired or wireless probes con nected to a tablet or mobile phone. Linear, curvilinear, and phasedarray probes are commonly available, and multifunctional probes are emerging. Linear high-

frequency probes have excellent image resolu tion but limited penetration, so they are used primarily to examine superficial structures. Deeper structures are visualized with curvilinear or phased-array probes, which have a lower frequency. Portable ultra sound devices offer two-dimensional or gray-scale imaging, and color flow and spectral Doppler imaging. Important considerations when purchasing an ultrasound machine include portability, image resolu tion, screen size, probe types, imaging modes, battery life, disinfection, image archiving capability, and warranty. COMMON APPLICATIONS ■ ■CARDIAC In the 1990s, clinicians began to perform focused cardiac POCUS exams to guide immediate management, especially for urgent and life-threatening conditions. In intensive care units and emergency departments, cardiac POCUS is routinely used to rapidly categorize shock states and acute respiratory failure. In outpatient settings, it is often used for serial monitoring of stable patients with chronic forms of heart disease. A limited or focused cardiac POCUS exam includes five core views: parasternal long-axis, parasternal short-axis (mid-ventricular or papil lary muscle level), apical four-chamber, subcostal four-chamber, and inferior vena cava views. Clinicians with comprehensive training in echocardiography, including cardiologists and intensivists certified by the National Board of Echocardiography, may perform advanced Doppler measurements of cardiac pressures and function. Cardiac POCUS and consultative echocardiography are complementary tech niques where the clinical situation and operator skill determine which approach is most appropriate. Cardiac POCUS exams can guide immediate and ongoing clinical decision-making when performed serially. To categorize shock states, left ventricular systolic function can be qualitatively categorized as nor mal (Video 493-1), hyperdynamic (Video 493-2), moderately reduced (Video 493-3), or severely reduced (Video 493-4). Other findings detected by cardiac POCUS that can change immediate management include acute right ventricular failure (Video 493-5), cardiac tampon ade (Video 493-6), and gross valvular abnormalities, including severe regurgitation of the tricuspid (Video 493-7), mitral (Video 493-8), and aortic (Video 493-9) valves, as well as large valvular vegetations (Video 493-10). Competence in basic cardiac POCUS has become a mandatory component of an increasing number of specialties, including emer gency medicine, pulmonary medicine, critical care medicine, and anesthesiology. ■ ■LUNG AND PLEURA Historically, thoracic ultrasonography, comprised of lung and pleural ultrasound, was established by clinicians specialized in critical care, pulmonary, and emergency medicine. The pleural surface can be imaged through the intercostal spaces using high-frequency probes, while low-frequency probes penetrate deeper, allowing visualization of structures in the thorax. Ultrasound is superior to chest x-ray for detection of pneumothorax, early interstitial processes, and small pleu ral effusions and is superior to chest CT for characterization of early complex pleural effusions. Pleural fluid is seen as a relatively hypoechoic space bounded by the diaphragm, chest wall, and atelectatic lung (Video 493-11). Pleural effusions are quantified as small (Video 493-12), moderate (Video 493-13), or large (Video 493-14), and qualitatively assessed as simple, homogeneously echogenic, complex nonseptated (Video 493-15), or complex septated (Video 493-16). Ultrasound guidance to identify an

Pleural effusion Dullness to percussion

4.8 0.1 Visualization of pleural fluid

0.07 Pulmonary edema Crackles 19–64 82–94 3.4 NS Bilateral B-lines

10.4 0.06 Pneumonia Bronchial breath sounds

3.3 NS Consolidation pattern 94–95 90–96 13.5 0.06 Elevated CVP (>8 cmH20) Neck vein inspection 47–92 93–96 9.7 0.3 CVP >10 mmHg (IVC size >2 cm)

4.9 0.32 Reduced ejection fraction 3rd heart sound (S3) 11–51 85–98 3.4 0.7 LV systolic dysfunction 84–91 85–88 6.5 0.14 FINDING SENSITIVITY (%) SPECIFICITY (%) LR+ LR– FINDING SENSITIVITY (%) SPECIFICITY (%) LR+ LR– Elevated LV pressure 4th heart sound (S4) 37–71 50–70 NS NS PCWP 17 if IVC >2.0

4.4 0.3 Pulmonary Cardiac PATHOLOGY PHYSICAL EXAMINATION POINT-OF-CARE ULTRASOUND PART 20 Emerging Topics in Clinical Medicine TABLE 493-1 Comparison of Physical Examination Versus Point-of-Care Ultrasound Findings for Common Pathologies Egophony 4–16 96–99 4.1 NS Decreased breath sounds

5.2 0.1 Crackles 19–67 36–94 1.8 0.8

Congestive heart failure Rales 12–23 88–96 NS NS Bilateral B-lines

19.4 0.03 Abdominojugular test 55–84 83–98 8.0 0.3 CVP >10 mmHg (IVC size >2 cm)

4.9 0.32 Ascites Bulging flanks 73–93 44–70 1.9 0.4 Visualized ascites

0.04 Urinary retention (>400 mL) Palpation

1.9 0.3 Bladder volume (>600 mL)

3.84 0.05 Lower extremity DVT Calf swelling >2 cm 61–67 69–71 2.1 0.5 Compression venous ultrasonography

0.04 Abbreviations: CVP, central venous pressure; DVT, deep-venous thrombosis; IVC, inferior vena cava; JVP, jugular venous pulse; LE, lower extremity; LR, likelihood ratio; LV, left ventricle or left ventricular; NA, not applicable; NS, not Abdomen Source: Reproduced with permission from A Bhagra et al: Point-of-care ultrasonography for primary care physicians and general internists. Mayo Clin Proc 91:1811, 2016. probability) 38–87 71–99 6.3 NA Elevated JVP 10–58 96–97 3.9 NS LE edema

93–96 NS NS Homan’s sign 10–54 39–89 NS NS Flank dullness 80–94 29–69 NS 0.3 Shifting dullness 60–87 56–90 2.3 0.4 Fluid wave 50–80 82–92 5.0 0.5 significant; PCWP, pulmonary capillary wedge pressure. Wells’ score (high Soft Tissue and Musculoskeletal

History & Physical Exam diagnostic procedure POCUS diagnose Consultative imaging Diagnosis Labs therapeutic procedure Treatment treat POCUS POCUS monitor No improvement Improvement POCUS screen Follow-up FIGURE 493-2 Clinicians can use point-of-care ultrasound (POCUS) as part of a patient’s diagnosis, treatment, monitoring, and screening. A patient encounter begins with the history and physical examination, followed by a focused bedside ultrasound exam to narrow the differential diagnosis and guide workup. Treatment plans can include bedside procedures that are performed with ultrasound guidance. Serial POCUS exams can monitor disease processes and guide ongoing treatment decisions. Screening POCUS exams can detect asymptomatic, potentially

treatable conditions. (Reproduced with permission from NJ Soni, BP Lucas: Diagnostic pointof-care ultrasound for hospitalists. J Hosp Med 10:120, 2015.) optimal site for pleural drainage reduces the risk of pneumothorax and bleeding complications. Normal air-filled lung tissue reflects sound waves, thereby pre venting visualization of aerated lung parenchyma. Two hallmarks of normal aeration of lung on ultrasound include lung sliding (Video 493-17), which results from respirophasic movement of the parietal and visceral pleural interface, and A-lines, which are horizontally ori entated reverberation artifacts seen deep to the pleural line of air-filled lungs (Video 493-18). Interstitial abnormalities manifest as B-lines, which are vertically orientated hyperechoic lines emanating from the pleural line to the bottom of the screen (Video 493-19). Depending on their density and distribution, B-lines can support a diagnosis of cardiogenic pulmonary edema, pneumonitis, acute respiratory distress syndrome, or interstitial lung diseases. Consolidation results in lung that is tissue dense on ultrasound. Mobile air bronchograms and blood flow detected by color flow Doppler are associated with pneumonia when seen in an area of consolidation (Video 493-20). Similar to chest x-ray and chest CT, identification of consolidation by lung ultrasound does not specify a diagnosis of pneumonia, and clinical correlation is required. ■ ■ABDOMEN Evaluation of peritoneal free fluid is a common abdominal POCUS application. POCUS cannot specify the type of fluid (i.e., ascites, blood, urine, bile, chyme) but can detect as little as 100–500 mL of peritoneal free fluid. When ascites is present (Video 493-21), POCUS can identify a safe site for paracentesis, improving procedural success and compli cation rates compared to landmark-based techniques. POCUS elimi nates attempts at paracentesis when an insufficient volume of ascites is present (Video 493-22). The best site, depth, and angle for needle insertion is determined using the ultrasound probe followed by color flow Doppler examination of the proposed trajectory of needle inser tion to avoid injury to abdominal wall blood vessels (Video 493-23). POCUS is used in the initial evaluation of acute renal failure and decreased urine output. Bladder ultrasound can rapidly identify presence or absence of urine in the bladder and confirm appropriate placement and function of a urinary catheter (Videos 493-24 and 493-25). Bladder ultrasound is more reliable than automated bladder

scanners for urinary retention, as bladder scanners can falsely report pelvic free fluid (i.e., ascites, cysts, small bowel obstruction) as elevated bladder volume. POCUS is effective to evaluate kidney size and echo genicity; identify renal cysts, large stones, and masses; and detect and grade hydronephrosis (Videos 493-26 to 493-28), thereby identifying obstructive uropathy.

POCUS can diagnose an abdominal aortic aneurysm (AAA) with high sensitivity and specificity (Videos 493-29 and 493-30). A pro tocol that emphasizes complete visualization of the abdominal aorta from celiac trunk through the iliac bifurcation in both transverse and longitudinal planes can provide a reliable evaluation of the aorta. POCUS use for AAA screening may reduce morbidity and mortality among high-risk patients. POCUS has utility for evaluation of small-bowel function. Normally, the small bowel is partially filled with air that obscures visualization due to scattering of sound waves. When a small-bowel obstruction (SBO) develops, the air-filled loops of bowel become fluid-filled, per mitting visualization of the bowel walls. Diagnostic criteria for SBO by ultrasound include dilation of the bowel (diameter >2.5 cm), fluidfilled small-bowel loops (confirmed by appearance of plicae circularis at the perimeter), and hyperactive to-and-fro peristalsis within loops of small bowel (Video 493-31). Combining patient history, physical examination, and a systematic survey of all four quadrants by ultra sound, clinicians can diagnose SBO rapidly and reliably. For a new diagnosis of SBO, POCUS can expedite early intervention and surgical consultation. For

recurrent SBO, POCUS can reduce repeat radiation exposure by CT scans and expedite initiation of medical management. CHAPTER 493 ■ ■LOWER EXTREMITY DEEP-VEIN THROMBOSIS Two-dimensional compression ultrasound is a rapid and accurate diag nostic technique for deep-vein thrombosis (DVT) that clinicians can learn after brief training programs. A point-of-care lower extremity com pression ultrasound exam yields similar diagnostic accuracy for detec tion of DVTs as traditional duplex or triplex ultrasound exams. DVTs commonly form at venous junctions because of high turbulence, and hence, compression ultrasound is performed at major branchpoints of the venous system. A perpendicular compression technique is required to ensure complete venous compression with wall-to-wall touching. A noncompressible vein is diagnostic of DVT (Video 493-32), and visual ization of intraluminal clot is not required to diagnose a DVT. Point-of-Care Ultrasound ■ ■SKIN AND SOFT TISSUE POCUS allows rapid differentiation between skin and soft tissue infec tions (SSTIs) and reactive lymph nodes, seromas, hematomas, hernias, thrombophlebitis, DVT, cysts, and bursitis. For SSTIs, POCUS can reduce unnecessary attempts at incision and drainage and avoid delays in surgical intervention. SSTIs range from cellulitis to phlegmon, abscess, and necrotizing fasciitis. POCUS can accurately distinguish abscess from cellulitis, but diagnostic accuracy is more variable for necrotizing fasciitis. To diagnose cellulitis, POCUS identifies subcuta neous edema described as “cobblestoning” (Video 493-33). Abscesses appear as irregular, enclosed areas superficially with compressible material and absent central flow on color Doppler (Video 493-34). ■ ■VASCULAR ACCESS Current evidence supports use of ultrasound guidance for insertion of central venous catheters (CVCs) in the femoral, internal jugular, and axillary veins. Ultrasound guidance for insertion of internal jugular CVCs improves procedure success rates and reduces complications, particularly pneumothorax and arterial punctures. A preprocedure ultrasound survey identifies potential vessels to cannulate and can reveal unsuspected venous thrombosis, atypical anatomy, and venous stenosis. During insertion, real-time visualization of the needle tip reduces procedure attempts and needle redirections, which reduces the risk of complications. Sonographic confirmation of the guidewire in the target vein provides a safety check prior to venous dilation and insertion of the CVC. For peripheral intravenous (PIV) catheter insertion, ultrasound can increase cannulation success rates while reducing puncture attempts, time to cannulation, and trauma to surrounding structures,

particularly in patients with anticipated difficult PIV placement or after failed attempts using standard techniques. Ultrasound identifies peripheral veins that are large, linear, and superficial, and real-time ultrasound guidance allows visualization of the needle tip entering the vessel lumen.

TRAINING ■ ■PATHWAYS Ultrasound training is a longitudinal process for clinicians as they progress through medical school, residency, and fellowship and enter clinical practice. Training recommendations for POCUS have been developed for different stages of medical education but with varying definitions of competence. Regardless of the clinical rank of the learner, competence in POCUS requires mastery of ultrasound knowledge (e.g., clinical indications, applications, limitations, artifacts), image acquisition, image interpretation, and clinical integration. Image acqui sition and interpretation skills are learned at varying rates and require deliberate practice. ■ ■CERTIFICATION Currently, there is no widely accepted certification for POCUS. Some residency and fellowship training programs, such as critical care and emergency medicine, require comprehensive training in POCUS, and hospitals generally grant POCUS privileges to physicians with board certification in these specialties. In contrast, internal medicine resi dency training does not require comprehensive POCUS training, and board certification in internal medicine does not

imply competence in POCUS. Several internal medicine residency programs and profes sional societies have developed POCUS training courses. PART 20 Emerging Topics in Clinical Medicine ■ ■CREDENTIALING AND PRIVILEGES Clinical privileges are governed by the rules and regulations of indi vidual hospitals. A hospital's credentialing and privileging committee is responsible for developing criteria for granting privileges for POCUS use, which may be guided by specialty-specific guidelines. Some hos pitals will designate a local POCUS expert to assess competence in POCUS prior to granting privileges for POCUS use in patient care. Hospital credentialing and privileging bodies may designate POCUS as a core privilege of a specialty (e.g., emergency medicine privileges include POCUS use) or as add-on privileges separate from the primary specialty's skills. Some well-established POCUS applications, such as ultrasound-guided CVC insertion, are commonly designated as core privileges when use of ultrasound guidance is standard of care. In contrast, less common POCUS applications, such as peripheral nerve blocks, may be designated as add-on privileges. FUTURE DIRECTIONS The increasing portability and affordability of ultrasound devices have allowed internal medicine clinicians to incorporate POCUS into front line patient care. Increasing POCUS use in internal medicine requires development of effective training programs during residency training and for internists in-practice. Tele-ultrasound has shown promise for training clinicians and delivering patient care remotely. In the coming years, artificial intelligence will facilitate both POCUS training and use in clinical care, and remote serial monitoring of common conditions like heart failure may be possible with patients' use of POCUS. ■ ■FURTHER READING American College of Emergency Physicians Ultrasound Guidelines: Emergency, point-of-care, and clinical ultrasound guidelines in medicine. Available at: https://www.acep.org/siteassets/ new-pdfs/policy-statements/ultrasound-guidelines--emergency-pointof-care-and-clinical-ultrasound-guidelines-in-medicine.pdf. Accessed December 3, 2024. Mayo PH et al: American College of Chest Physicians/La Societe de Reanimation de Langue Francaise statement on competence in criti cal care ultrasonography. Chest 135:1050, 2009. Qaseem A et al: Appropriate use of point-of-care ultrasonography in patients with acute dyspnea in emergency department or inpatient

settings: A clinical guideline from the American College of Physicians. Ann Intern Med 174:985, 2021. Soni NJ et al: Point-of-care ultrasound for hospitalists: A position statement of the Society of Hospital Medicine. J Hosp Med 14:E1, 2019. Soni NJ, Arntfield R, Kory PD: Point-of-Care Ultrasound, 2nd ed. Philadelphia, Elsevier/Saunders, 2019. Spencer KT et al: Focused cardiac ultrasound: recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 26:567, 2013. VIDEO 493-1 Normal cardiac function. VIDEO 493-2 Hyperdynamic cardiac function. VIDEO 493-3 Reduced cardiac function. VIDEO 493-4 Severely reduced cardiac function. VIDEO 493-5 Acute right heart failure. VIDEO 493-6 Cardiac tamponade. VIDEO 493-7 Tricuspid valve regurgitation. VIDEO 493-8 Mitral valve regurgitation. VIDEO 493-9 Aortic valve regurgitation. VIDEO 493-10 Tricuspid valve vegetation. VIDEO 493-11 Lung atelectasis. VIDEO 493-12 Small pleural effusion. VIDEO 493-13 Moderate pleural effusion. VIDEO 493-14 Large pleural effusion. VIDEO 493-15 Homogenous pleural effusion. VIDEO 493-16 Loculated pleural effusion. VIDEO 493-17 Pleural sliding. VIDEO 493-18 A-Line artifact. VIDEO 493-19 B-Line artifact. VIDEO 493-20 Lung consolidation. VIDEO 493-21 Large-volume ascites. VIDEO 493-22 Small-volume ascites. VIDEO 493-23 Abdominal wall vessels with color Doppler. VIDEO 493-24 Urinary catheter balloon in empty bladder. VIDEO 493-25 Malfunctioning urinary catheter. VIDEO 493-26 Mild hydronephrosis. VIDEO 493-27 Moderate hydronephrosis. VIDEO 493-28 Severe hydronephrosis. VIDEO 493-29 Abdominal aortic aneurysm. VIDEO 493-30 Abdominal aortic aneurysm. VIDEO 493-31 Small-bowel obstruction. VIDEO 493-32 Deep-vein thrombosis. VIDEO 493-33 Cellulitis. VIDEO 493-34 Simple

abscess.

03 - 495 Complementary and Integrative Therapies and Practices

495 Complementary and Integrative Therapies and Practices

seen as intractable, even the most effective drugs will not work if physi cians fail to prescribe them and if patients fail to take them. Although the dominant forms of investigation in medicine seek cellular or molecular therapeutic targets to modify disease, behavioral sciences have revealed cognitive pathways that operate nearly as predictably as the genetic code. The opportunity for behavioral economics to improve health and health care delivery derives from its recognition of these behavioral pathways and the growing empirical evidence about how to best make use of them.

■ ■FURTHER READING Asch DA et al: Automated hovering in health care—Watching over the 5000 hours. N Engl J Med 367:1, 2012. Chater N, Loewenstein G: The i-frame and the s-frame: How focus ing on individual-level solutions has led behavioral public policy astray. Behav Brain Sci 46:e147, 2023. Loewenstein G et al: Asymmetric paternalism to improve health behaviors. JAMA 298:2415, 2007. Thaler RH et al: Choice architecture, in The Behavioral Foundations of Public Policy, E. Shafir (ed). Princeton, NJ, Princeton University Press, 2013, pp 428–439. Volpp KG et al: Financial incentive-based approaches for weight loss: A randomized trial. JAMA 300:2631, 2008. Helene M. Langevin

Complementary and

Integrative Therapies

and Practices PART 20 Emerging Topics in Clinical Medicine The search for health and improved well-being includes many treatments, practices, and systems of care that may have originated outside conven tional medicine but are gradually being folded into mainstream health care. The

current health care system is fragmented, often emphasizing the pharmacologic treatment of disease alone, while often neglect ing the promotion, support, and, importantly, restoration of health. Though the disease-focused model is dominant in our research and health care ecosystem, there has been a longstanding awareness that many chronic diseases, including pain conditions, can be prevented or better managed by incorporating nonpharmacologic interventions such as nutrition, exercise, and stress management into care, with an emphasis on understanding the person as a whole. Many complemen tary practices follow this model, and there is preliminary evidence indicating that these approaches lead to improved self-care, a better personal sense of well-being, and a greater commitment to a healthy lifestyle. Integrative health emphasizes not only the integration of complementary and conventional care but also an integrative approach to treatment of the whole person. This includes expanding our under standing of how physiologic systems interact with one another and of the connections between physical, psychological, and social aspects of health. Integrative health also includes striving for a better understand ing of “salutogenesis” or pathogenesis in reverse, meaning the process by which health is restored when recovering from an injury, acute ill ness, or the exacerbation of a chronic disease, or when a “predisease” condition such as prediabetes or prehypertension is reversed through changes in behavior rather than pharmacologic treatment. DEFINITIONS AND SCOPE Complementary health therapies and practices include a broad range of practices, interventions, and natural products that are not typically part of conventional medical care (Table 495-1). The term complementary

refers to the use of these practices together with conventional therapies and is increasingly preferred to the term alternative, which denotes usage as a substitute for standard care. The term integrative health care refers to conventional and comple mentary therapies and practices used together in a coordinated way. Integrative health also emphasizes care of the whole person that aims to improve health in multiple interconnected domains: social, psychologi cal, and physical, including multiple organs and systems. The term whole person health involves looking at the whole person— not just separate organs or body systems—and considering multiple factors that promote either health or disease. It means helping and empowering individuals, families, communities, and populations to improve their health in multiple interconnected biological, behavioral, social, and environmental areas. Instead of treating a specific disease, whole person health focuses on restoring health, promoting resilience, and preventing diseases across a lifespan. The use of integrative approaches to health and well-being has grown within care settings across the United States. Researchers are currently exploring the potential benefits of integrative health in a variety of situations, including pain management for military person nel and veterans, relief of symptoms in cancer patients and survivors, and programs to promote healthy behaviors. Although complementary therapies and practices vary widely, it is useful to classify them by their primary therapeutic input, which may be dietary (e.g., diet, herbs), psychological (e.g., meditation), physical (e.g., massage, acupuncture), or the combination of psychological and physi cal (e.g., yoga, tai chi). Although some complementary health practices are recommended or provided by a physician or a complementary health care provider such as a chiropractor, acupuncturist, or naturo pathic practitioner, many of these practices are undertaken as “selfcare.” Although some are reimbursed, most are paid for out of pocket. PATTERNS OF USE The first large survey of use of complementary health practices was performed by David Eisenberg and associates in 1993. It surprised the medical community by showing that >30% of Americans use complementary health products and practices. Many surveys since that time have extended those conclusions. The National Health Interview Survey (NHIS), a large, national household survey in which thousands of

Americans are interviewed about their health- and illness-related experiences, is conducted annually by the National Center for Health Statistics, a component of the Centers for Disease Control and Preven tion. This survey, which addressed the use of complementary health practices in 2002, 2007, 2012, 2017, and 2022 uses methods that create a nationally representative sample and has a sample size large enough to permit valid estimates about some subgroups. An analysis of data from 27,651 adults in the most recent survey, which was conducted in 2022, evaluated changes in the U.S. adult use of seven complementary health therapies and practices over a 20-year period (from 2002 to 2022): yoga, meditation, massage therapy, chiro practic care, acupuncture, naturopathy, and guided imagery/progressive muscle relaxation. Over 20 years, U.S. adults not only increased their overall use of complementary health approaches but were also more likely to use them specifically for managing pain. In 2022, 36.7% of people used at least one of the seven approaches, compared to 19.2% in 2002. Use of yoga, meditation, and massage therapy increased the most from 2002 to 2022. Use of yoga increased from 5% in 2002 to 15.8% in 2022, rising from the fifth to the second most-used practice. Medita tion increased from 7.5% in 2002 to 17.3% in 2022, and it remained the most-used complementary health practice over the 20 years. The 2012 survey, for which there are data about use of natural products, yielded the estimate that nonvitamin, nonmineral dietary supplements are used by ~18% of adults and 5% of children. Americans often pay out-of-pocket for complementary health products and practices; the estimated out-of-pocket expenditure for complementary health practices in 2012 was $30.2 billion ($28.3 bil lion for adults and $1.9 billion for children), representing 1.1% of total health expenditures and 9.2% of out-of-pocket costs. On visits to com plementary practitioners, Americans spent $14.7 billion out of pocket,

TABLE 495-1 Glossary of Complementary and Integrative Health Therapies and Practices Acupuncture A family of procedures involving stimulation of defined anatomic points, a component of the major Asian medical traditions; most common application involves penetrating the skin with thin, solid, metallic needles that are manipulated by the hands or by electrical stimulation Ayurvedic medicine The major East Indian traditional medicine system; treatment combines products (mainly derived from plants, but may also include animal, metal, and mineral), diet, exercise, and lifestyle Biofeedback The use of electronic devices to help people learn to consciously control body functions such as breathing or heart rate Chiropractic Chiropractic care involves the adjustment of the spine and joints to influence the body’s nervous system and natural defense mechanisms to alleviate pain and improve general health; primarily used to treat back problems, headaches, nerve inflammation, muscle spasms, and other injuries and traumas Dietary supplement A product that is intended to supplement the diet, is taken by mouth, contains one or more dietary ingredients (including vitamins, minerals, herbs, amino acids, or certain other substances), and is labeled as being a dietary supplement Homeopathy A medical system with origins in Germany that is based on a core belief in the theory of “like cures like”—compounds that produce certain syndromes, if administered in very diluted solutions, will be curative Hypnosis The induction of an altered state of consciousness characterized by increased responsiveness to suggestion Massage Manual therapies that manipulate muscle and connective tissues to enhance the function of those tissues and promote muscle relaxation and well-being Meditation A group of practices, largely based in Eastern spiritual traditions, intended to focus or control attention and obtain greater awareness of the present moment, or mindfulness Mind and body practices A large and diverse group of procedures or techniques that are administered or taught by a trained practitioner or teacher; examples include acupuncture, massage therapy, meditation, relaxation

techniques, spinal manipulation, tai chi, and yoga Natural products A variety of products such as herbs (also known as botanicals), vitamins and minerals, and probiotics, which are widely marketed, readily available to consumers, and often sold as dietary supplements Naturopathy A clinical discipline that emphasizes a holistic approach to the patient, herbal medications, diet, and exercise; practitioners have degrees as doctors of naturopathy Osteopathy A clinical discipline, now incorporated into mainstream medicine, that historically emphasized spinal manipulative techniques to relieve pain, restore function, and promote overall health Qigong A mind and body practice originating in China that involves using exercises to optimize energy within the body, mind, and spirit, with the goal of improving and maintaining health and well-being Relaxation techniques A number of practices such as progressive relaxation, guided imagery, biofeedback, self-hypnosis, and deep breathing exercises, with the goal of producing the body's natural relaxation response, characterized by slower breathing, lower blood pressure, and a feeling of increased well-being Spinal manipulation, osteopathic manipulation A technique where practitioners use their hands or a device to apply a controlled thrust (i.e., a force of a specific magnitude or degree in a specific direction) to a joint of the spine Tai chi A mind and body practice originating in China that involves slow, gentle movements and sometimes is described as "moving meditation" Traditional Chinese medicine A medical system that uses acupuncture, herbal mixtures, massage, exercise, and diet which is almost 30% of what they spent out of pocket on services by conventional physicians ($49.6 billion). On natural products, such as dietary supplements, Americans spent $12.8 billion out of pocket, which was about one-quarter (24%) of what they spent out of pocket on prescription drugs ($54.1 billion). Trends are even more striking for pain conditions. According to the NHIS surveys, painful conditions are the most common reasons why American adults use complementary health products and practices. About 40 million American adults experience severe pain in any given year, and they spend >$14 billion out of pocket on complementary therapies to manage their pain. A recent analysis of NHIS data showed a notable rise in the proportion of U.S. adults using complementary health approaches specifically for pain management. Among participants using any of the complementary health approaches, the percentage reporting use for pain management increased from 42.3% in 2002 to 49.2% in 2022. Some patients seek out complementary health practitioners because they offer greater personal attention. For others, therapies and prac tices perceived as outside the mainstream reflect a "self-help" approach to health and well-being or satisfy a search for "natural" or less invasive alternatives. Since dietary supplements are labeled as "natural," they are often believed, incorrectly, to be inherently healthy. CATEGORIES OF COMPLEMENTARY AND INTEGRATIVE HEALTH THERAPIES AND PRACTICES BASED ON PRIMARY THERAPEUTIC INPUT ■ ■PRIMARY DIETARY INPUT Natural products, including plant and animal products, have a long and impressive history as sources of medicine and as important resources

for biologic research. Whether as herbal supplements or as part of a diet, natural products are frequently consumed as a complex mixture of substances. This complexity can be further amplified by potential interactions with endogenous metabolic pathways, including those associated with the microbiome. The result is a collection of natural products and their metabolites that, individually and/or collectively, are associated with a network of biologic activity. Importantly, in addi tion to direct action on biologic targets, the activity of natural products can be influenced by an individual's diet, health, and metagenomic background. Although much remains to be understood about mecha nisms of action, results of research on some natural products for a few conditions

appear promising. In addition, in the 2012 NHIS, users of natural product supplements were twice as likely to report taking the natural product for a general well-being reason than for treatment of a specific health condition (88.9 vs 44.9%, respectively). Although to date, research on natural products has focused on their use for specific diseases as outlined below, a better understanding is needed about how natural products, including food, can be used most effectively to sup port health. Cannabinoids An increasing amount of attention has been given recently to the nonpsychogenic effects of cannabinoids, such as can nabidiol (CBD), and terpenes found in the cannabis plant on chronic pain, particularly neuropathic pain; studies have found some limited evidence that these medicines produced better pain relief than place bos. Cannabinoids (cannabis extract, synthetic tetrahydrocannabinol [THC]) have been studied for therapeutic effects in multiple sclerosis (MS) and may relieve spasticity as well as pain in people with MS; how ever, no marijuana-derived medications are approved by the U.S. Food and Drug Administration (FDA) to treat MS. Sativex, an oral mucosal spray containing a mixture of THC and CBD, has received regulatory

approval in >25 countries outside the United States for the treatment of spasticity (muscle stiffness/spasm) due to MS. Sativex is currently licensed in the United Kingdom for use as an add-on treatment for MS-related spasticity when people have shown inadequate response to other symptomatic treatments. Importantly, the psychoactive prop erties and other potential adverse effects of preparations containing cannabinoids need to be considered, including interactions with other medications and natural products; more research is needed in this area.

Melatonin Melatonin has been shown to help reduce anxiety in patients who are about to have surgery and may be as effective as standard treatment with midazolam in reducing preoperative anxiety. Findings from clinical trials support the use of melatonin supplements for sleep problems caused by shift work or jet lag and for improving sleep-onset latency and daytime sleepiness in people with insomnia. However, there are safety concerns about the use of melatonin by chil dren and teenagers. U.S. sales of melatonin increased by about 150% during the COVID-19 pandemic, and the number of reports to U.S. poison control centers about pediatric melatonin ingestion increased from 8337 in 2012 to 52,563 in 2021. Further, according to a study pub lished in JAMA, a majority of melatonin “gummy” products were inac curately labeled, with most products exceeding the declared amount of melatonin and CBD. Omega-3 Fatty Acids Clinical trials on rheumatoid arthritis (RA) have found that fish oil supplements can help alleviate tender joints and morning stiffness and reduce the daily nonsteroidal anti-inflammatory drug (NSAID) requirement of RA patients; however, data are not as definitive for other pain conditions. Gamma-linolenic acid (GLA) is an omega-6 fatty acid found in the oils from some plants, including evening primrose (Oenothera biennis), borage (Borago officinalis), and black currant (Ribes nigrum). Although oils containing GLA may have some benefit in relieving RA symptoms, only a few studies have been conducted on each of the oils. At present, it is uncertain whether omega-3 fatty acid supplementation is useful for depression. Some studies have shown small effects in adjunctive therapy in patients with a diagnosis of major depressive disorder (MDD) and in depres sive patients without a diagnosis of MDD; however, most trials have been adjunctive studies. Controlled trials of omega-3 fatty acids as monotherapy are inconclusive compared to standard antidepressant medicines, and it remains unclear whether a mechanism is present to suggest that a pharmacologic or biologic antidepressant effect exists. Furthermore, there is evidence that a high dosage of fish oil supple mentation is associated with a significant increased risk of atrial fibril lation (AF) compared with placebo. PART 20 Emerging Topics in Clinical Medicine Antioxidants Findings from the Age-Related

Eye Disease Studies (AREDS and AREDS2) suggest that dietary supplementation with antioxidant vitamins may slow the progression of age-related macular degeneration (AMD). Compared to the original AREDS formulation, the AREDS2 formulation replaced beta-carotene with lutein and zea xanthin due to increased risk of cancer in smokers taking high dose beta-carotene. Of note, in AREDS2, supplementation with lutein/ zeaxanthin only appeared to be beneficial in participants with low dietary lutein and zeaxanthin. In a similar vein, a study using baseline data from the AREDS cohort reported that individuals eating healthier diets, characterized by higher intake of vegetables, whole grains, and seafoods, compared to those eating less healthy “Western” diets, were less likely to show signs of early AMD. It is therefore unclear at present whether the AREDS formula should be recommended for the general population regardless of diet. Challenges of Research on Natural Products One challenge in this area is the extremely varied doses of natural products that are sold over the counter and used without much guidance or evidence of effi cacy. We also know from research on vitamins that “more is not neces sarily better” and that taking a “natural” substance such as a vitamin in quantities that greatly exceed what is found in food can be harmful. Additional challenges in the assessment of plant products include their complexity and variability, including possible instability of active

components or the presence of impurities, conflicting or unreliable conclusions in the literature, and low statistical power of studies. Fur ther, there is a paucity of data on the safety of many products, including the safety of their use in a twenty-first-century context (e.g., if taken with modern prescription drugs) and their appropriate use in the con text of traditional or indigenous practices. Regulation There is an important distinction between natural products sold as dietary supplements and drugs developed from natural sources that are used to treat specific diseases. The Dietary Supplement Health and Education Act (DSHEA), passed in 1994, gives authority to the FDA to regulate dietary supplements, but with expecta tions that differ in many respects from the regulation of drugs or food additives. Purveyors of dietary supplements cannot claim that they prevent or treat any disease. They can, however, claim that they main tain “normal structure and function” of body systems. For example, a product cannot claim to treat arthritis, but it can claim to maintain “normal joint health.” Homeopathic products predate FDA drug regulations and are sold with no requirement that they be proved effective. Although homeo pathic products are widely believed to be safe because they are highly dilute, one product, a nasal spray called Zicam, was withdrawn from the market when it was found to produce anosmia, probably because of significant zinc content. In January 2017, the FDA warned consumers about homeopathic teething tablets containing belladonna that pose a serious risk to infants and children. Regulation of advertising and marketing claims is the purview of the Federal Trade Commission (FTC). The FTC does take legal action against promoters or websites that advertise or sell dietary supplements with false or deceptive statements. Misleading marketing of dietary supplements, homeopathic products, and indeed other complementary health products and practices contributes to the very significant risk that individuals will use them instead of effective conventional modali ties. For example, in April 2020, the FTC sent warning letters to several companies allegedly selling unapproved products—some of which included high-dose dietary supplements—that may violate federal law by making deceptive or scientifically unsupported claims about their ability to treat or cure COVID-19. Inherent Toxicity Although the public may believe that “natu ral” equates with “safe,” it is abundantly clear that natural products can be toxic. Misidentification of medicinal mushrooms has led to liver failure. Contamination of tryptophan supplements caused the eosinophilia-myalgia syndrome. Herbal products containing particular species of Aristolochia were associated with

genitourinary malignan cies and interstitial nephritis. In 2013, dietary supplements containing 1,3-dimethylamylamine (DMAA), often touted as a “natural” stimu lant, led to cardiovascular problems, including heart attacks. Among the most controversial dietary supplements is Ephedra sinica, or ma huang, a product used in traditional Chinese medicine for short-term treatment of asthma and bronchial congestion. The scientific basis for these indications was revealed when ephedra was shown to contain ephedrine alkaloids, especially ephedrine and pseudoephedrine. With the promulgation of the DSHEA regulations, supplements containing ephedra and herbs rich in caffeine sold widely in the U.S. marketplace because of their claims to promote weight loss and enhance athletic performance. Reports of severe and fatal adverse events associated with use of ephedra-containing products led to an evidence-based review of the data surrounding them, and in 2004, the FDA banned their sale in the United States. A major current concern with dietary supplements is adulteration with pharmacologically active compounds. Multi-ingredient products marketed for weight loss, bodybuilding, “sexual health,” and athletic performance are of particular concern. Recent FDA recalls have involved contamination with steroids, diuretics, stimulants, and phos phodiesterase type 5 inhibitors. Herb-Drug Interactions A number of natural products have potential impacts on the metabolism of drugs. This effect was illustrated most compellingly with the demonstration in 2000 that consumption

TABLE 495-2 Resources for Dietary Supplement–Drug Interactions National Institutes of Health National Center for Complementary and Integrative Health (NCCIH) https://www.nccih.nih.gov/health/know-science/how-medications-supplements-

interact The National Institutes of Health NCCIH Know the Science initiative provides information for patients about complex scientific health topics such as drugsupplement interactions. Medscape http://www.medscape.com/druginfo/druginterchecker?cid=med This website is maintained by WebMD and includes a free drug interaction checker tool that provides information on interactions between two or more drugs, herbals, and/or dietary supplements. NatMed https://naturalmedicines.therapeuticresearch.com/tools/interaction-checker.aspx This website provides an interactive natural product–drug interaction checker tool that identifies interactions between drugs and natural products, including herbals and dietary supplements. This service is available by subscription. of St. John’s wort interferes with the bioavailability of the HIV protease inhibitor indinavir. Later studies showed its similar interference with metabolism of topoisomerase inhibitors such as irinotecan and with cyclosporine and many other drugs. The breadth of interference stems from the ability of hyperforin in St. John’s wort to upregulate expression of the pregnane X receptor, a promiscuous nuclear regulatory factor that promotes the expression of many hepatic oxidative, conjugative, and efflux enzymes involved in drug and food metabolism. Because of the large number of compounds that alter drug metabo lism and the large number of agents some patients are taking, identifi cation of all potential interactions can be a daunting task. Several useful Web resources are available as information sources (Table 495-2). Clearly, attention to this problem is particularly important with drugs with a narrow therapeutic index, such as anticoagulants, antiseizure medications, antibiotics, immunosuppressants, and cancer chemo therapeutic agents. Although there are many examples of substances of natural origin successfully used as pharmaceutical drugs, in general, natural products ingested as food or herbal teas, rather than concen trated extracts, are less likely to cause harm. ■ ■PRIMARY PSYCHOLOGICAL INPUT Therapies and practices whose primary therapeutic input is predomi nantly mental include conventional types of psychotherapy, such as cognitive behavioral therapy (CBT), and

complementary practices, such as meditation and mindfulness-based stress reduction (MBSR). Relaxation techniques, including biofeedback-assisted relaxation, also fall into this category. The boundary between conventional and com plementary can be blurred, as CBT programs, for example, frequently incorporate elements of MBSR and relaxation techniques. These therapies and practices are being gradually integrated into aspects of conventional care, such as cardiac rehabilitation programs, and are playing an increasingly recognized role in the management of pain, as well as stress and sleep disturbances. Cognitive Behavioral Therapy (CBT) The American College of Physicians practice guidelines (2016) strongly recommend the use of CBT for insomnia (also called CBT-I) as the initial treatment for chronic insomnia. Although CBT-I often includes relaxation tech niques, it is not clear whether relaxation alone is beneficial. Various online applications are increasing the accessibility of these techniques at low cost. Mindfulness-Based Stress Reduction (MBSR) Mindfulness meditation has been found to significantly reduce pain in experimen tal and clinical settings and to improve a wide spectrum of clinically relevant cognitive and health outcomes, including low-back pain and fibromyalgia. Recent findings from neuroimaging and randomized controlled trials confirm that mindfulness meditation reduces pain by engaging multiple, unique, nonopioidergic mechanisms that are

distinct from placebo and that vary across meditative training level. There is some growing evidence that mindfulness meditation can have a beneficial effect on anxiety and help people recover from substance use disorders.

Hypnosis Findings from a few studies have demonstrated that train ing patients in the use of self-hypnosis significantly reduced their need for sedatives and analgesia when undergoing interventional radiologic procedures. Some studies also have suggested that hypnosis may be help ful for anxiety and health-related quality of life in people with irritable bowel syndrome (IBS). There is some evidence to suggest that hypno therapy may improve smoking cessation, but data are not definitive. Relaxation Techniques Relaxation techniques, including biofeed back and progressive muscle relaxation, may be helpful in managing a variety of stress-related health conditions, including anxiety associated with ongoing health problems and in those who are having medical procedures. Diaphragmatic breathing exercises may modestly lower blood pressure, reduce levels of cortisol, and reduce glycemia in people with type 2 diabetes. The efficacy of biofeedback has been evaluated in numerous studies for tension headaches, with positive results. Several studies have shown that biofeedback decreased the frequency of both pediatric and adult migraines, with some showing an effect lasting over an average follow-up phase of 17 months. Evidence suggests that relaxation techniques may also provide some benefit for symptoms of posttraumatic stress disorder and help reduce occupational stress in health care workers. Clinical practice guidelines issued by the Ameri can Cancer Society on the evidence-based use of integrative therapies during and after breast cancer treatment recommend yoga for anxiety and stress reduction. For some of these conditions, relaxation tech niques are used as an adjunct to other forms of treatment.
CHAPTER 495 ■ ■PRIMARY PHYSICAL INPUT A physical therapeutic input can be delivered manually (e.g., massage) or using a device (e.g., acupuncture) or can be generated by the patient (e.g., exercise). Complementary and Integrative Therapies and Practices
Acupuncture The role of acupuncture in pain management has been controversial for decades, with critics pointing out its “prescien tific” theoretical basis, and indeed, the rationale for the use of specific “acupuncture points” remains to be established. However, recent largescale meta-analyses have demonstrated acupuncture to be superior to both usual care and sham acupuncture for

chronic musculoskeletal pain, headache, and osteoarthritis (OA), with beneficial treatment effects persisting for up to 12 months. Clinical practice guidelines issued by the American College of Rheumatology and the Arthritis Foundation conditionally recommend acupuncture for knee, hip, and/ or hand OA. The most recent (2017) American College of Physicians clinical guidelines recommend acupuncture as one of the initial treat ment options for patients with acute, subacute, and chronic low-back pain. Acupuncture may provide a modest reduction in symptoms of depression, particularly when compared with no treatment or a con trol. Acupuncture or electroacupuncture may be an appropriate addi tion to drug treatment for managing chemotherapy-induced nausea and vomiting in patients with cancer. Clinical guidelines issued by the Society for Integrative Oncology and the American Society of Clinical Oncology in 2022 found intermediate level of evidence (with moderate strength) to recommend that acupuncture, reflexology, acupressure, or massage may help relieve pain in people with cancer. Acupuncture may relieve symptoms of allergic rhinitis. Clinical practice guidelines from the American Academy of Otolaryngology–Head and Neck Surgery include acupuncture among the options that health care providers may offer to interested patients with allergic rhinitis. Spinal Manipulation The role of both osteopathic and chiroprac tic spinal manipulative therapies (SMTs) in management of low-back pain also has been the subject of a number of carefully performed tri als and many systematic reviews. Conclusions are not consistent, but the American College of Physicians guidelines conclude that spinal manipulation has a small effect on improving function and pain com pared with control—either a sham manipulation or an inert treatment.

Although evidence for spinal manipulation for chronic low-back pain is graded as low quality, the recommendation for consideration of nonpharmacologic treatment including spinal manipulation is graded as a strong recommendation, reflecting increasing concern with the impact of chronic opioid use for low-back pain. The evidence of benefit of spinal manipulation for neck pain is not as extensive, and continued concern that cervical manipulation may occasionally precipitate vascu lar injury clouds a contentious debate.

Massage Low- to moderate-quality evidence suggests that massage therapy is superior to nonactive therapies in reducing arthritis pain and improving functional outcomes. Massage may provide short-term relief from low-back pain, but the evidence is not of high quality. There is some evidence that massage has a positive effect on migraine, tension headaches, and neck pain. ■ ■COMBINED PSYCHOLOGICAL AND

PHYSICAL INPUT The primary therapeutic input for other mind and body practices is a combination of physical and psychological. Examples of practices in this category include yoga and tai chi, which combine movement, physical postures, and meditation. Yoga Yoga can be beneficial for patients with fibromyalgia or chronic low-back pain, and yoga compared to nonexercise controls results in small to moderate improvements in back-related function at 3 and 6 months. There is overall evidence that yoga benefits people’s general well-being by relieving stress, supporting good health habits, and improving mental/emotional health and sleep. Yoga can also help with quitting smoking, anxiety or depressive symptoms associated with dif ficult life situations, and quality of life for people with chronic diseases. PART 20 Emerging Topics in Clinical Medicine Tai Chi Clinical practice guidelines issued by the American College of Rheumatology and the Arthritis Foundation strongly recommend tai chi, along with other nondrug approaches such as self-management programs, for managing knee and/or hip OA. Tai chi has been shown to improve overall motor

function, including balance and stability in older adults. Tai chi may help improve sleep quality in individuals with mild insomnia. Tai chi also has been shown to improve quality of life in people with heart disease, cancer, and other chronic illnesses. MULTICOMPONENT THERAPIES

AND SYSTEMS Multicomponent approaches to health comprise two or more interven tions such as lifestyle changes, physical rehabilitation, psychotherapy complementary health practices, and conventional medicine in various combinations, with an emphasis on whole person health. Complemen tary health therapies and practices are often multicomponent in nature, both in traditional health systems (e.g., traditional Chinese medicine, naturopathy) and in modern integrative practice. The U.S. Veterans Health Administration uses a multicomponent model of pain care that emphasizes nonpharmacologic methods, both conventional (e.g., physi cal therapy, CBT) and complementary (e.g., yoga, acupuncture), and may also include nutrition consultations. Several medical systems, such as chiropractic, osteopathy, naturopathy, and homeopathy, that arose in the late nineteenth century continue to be practiced today. Osteopathic medicine is mostly integrated into conventional medicine, with the addi tion of specific osteopathic musculoskeletal manipulation techniques. While homeopathy and naturopathy have remained largely separate from mainstream medicine, chiropractic care is increasingly available in some conventional care settings. A number of multicomponent systems, often called “whole health” systems, such as traditional Chinese medicine, Ayurveda, and homeopathy, use a diagnostic and therapeutic framework that is different from that of conventional medicine, which has posed additional challenges to their rigorous investigation. Naturopathy Naturopathy, or naturopathic medicine, is a multi component therapeutic system based on philosophical principles that guide practice. Naturopaths prescribe conventional and unconven tional diagnostic tests and medications, with an emphasis on relatively low doses of drugs, herbal medicines, healthy diet, and exercise.

Chiropractic The practice of chiropractic care, founded by David Palmer in 1895, is the most widespread practitioner-based comple mentary health practice in the United States. Although the scope of practice varies widely, chiropractic practice emphasizes manual thera pies for treatment of musculoskeletal complaints. Osteopathic Medicine Founded in 1892 by the physician Andrew Taylor Still, osteopathic medicine was originally based on the belief that manipulation of soft tissue and bone can correct a wide range of diseases of the musculoskeletal and other organ systems. Over the ensuing century, the osteopathic profession has welcomed increasing integration with conventional medicine. Today, the postgraduate train ing, practice, credentialing, and licensure of osteopathic physicians are virtually indistinguishable from those of allopathic physicians. Osteo pathic medical schools, however, include training in manual therapies, particularly spinal manipulation, as well as diagnostic methods based on palpation of musculoskeletal tissues that are not part of conven tional medical education. Homeopathy The theoretical framework of homeopathy is based on two unconventional principles: “like cures like,” the notion that a disease can be cured by a substance that produces similar symptoms in healthy people; and the “law of minimum dose,” the notion that the lower the dose of the medication, the greater its effectiveness. Although the current lack of biologic underpinning for these prin ciples has seriously limited the rationale for their use, the diagnostic framework of homeopathy could be the source of new insights that could be explored. As previously discussed, the regulatory framework for homeopathic remedies differs from that for dietary supplements, in that homeopathic products are regulated as drugs under the Federal Food, Drug, and Cosmetic Act and are subject to the same require ments related to approval, adulteration, and misbranding as other drug products. There are currently no

homeopathic products approved by the FDA. Homeopathic remedies are widely available and commonly recommended by naturopathic physicians, chiropractors, and other licensed and unlicensed practitioners. Challenges of Clinical Research on Multicomponent Therapies and Systems Classic randomized controlled trial (RCT) designs may not be well suited for research on multicomponent complemen tary interventions and systems such as naturopathy and Ayurvedic medicine. The dynamic relationships among an array of factors that affect health and well-being is inherent to the philosophy of these systems of care and poses methodologic challenges to the effec tive application of conventional RCT design. Pragmatic comparative effectiveness designs with “usual care” comparators are widely used to study these types of interventions, and trials may need to take into account the individualization of interventions and the underlying theories of these multicomponent systems. Thus, a key component of research on multicomponent therapeutic systems is the development of validated and reproducible “manualized” treatment protocols allowing for some flexibility and individual patient care. Pragmatic studies that compare multicomponent treatments with usual care cannot determine which treatment components are responsible for benefits, but other kinds of translational studies can address this issue. THERAPEUTIC OUTPUT—SYSTEMS IMPACTED AND CHALLENGES OF MECHANISTIC RESEARCH Complementary and integrative interventions whose therapeutic input is dietary, psychological, and/or physical may exert their effects, or therapeutic output, through a variety of mechanisms and physiologic systems. For example, peppermint oil may relieve pain associated with IBS by directly relaxing gastrointestinal smooth muscle, probiotics may have effects on the nervous system as well as the gut, and some components of traditional Chinese medicine, as well as omega-3 fatty acids and their derivatives, have immune-mediated anti-inflammatory effects. Multicomponent interventions with psychological and/or phys ical therapeutic input such as meditation and acupuncture can have effects on the nervous system and may also target other body systems

affected by the pain condition; for example, tai chi may improve bal ance and stability by increasing flexibility and core strength, and the stretching involved in yoga may improve low-back pain by reducing connective tissue inflammation. For all types of therapeutic input, biopsychosocial interactions also may be important; for example, par ticipation in an integrative group therapy pain management program may provide tools to help relieve symptoms of anxiety and depression as well as pain. Deepening the scientific understanding of the connections that exist across domains of human health is important to better understand how conditions interrelate, identify multicomponent interventions that address these problems, and increase the support of patients through the full continuum of their health experience, including the return to health. Studies of multicomponent interventions often require multi disciplinary expertise and use state-of-the-art techniques in areas such as neuroscience, immunology, pharmacognosy, proteomics, genetics, and epigenomics. Further, there are limited preclinical models for some complementary health interventions (e.g., no relevant animal model for meditative movement practices such as yoga or tai chi). Objective, validated measurement tools are essential, as are processes and proce dures to ensure quality control, whether the intervention is a mind and body practice or a natural product. PATIENT AND PROVIDER RESOURCES Physicians regularly face difficult challenges in providing patients with advice and education about complementary health therapies and prac tices. Of particular concern to all physicians are practices of uncertain safety and practices that raise inappropriate hopes. Cancer therapies, antiaging regimens, weight-loss programs, and products that claim to improve sexual function or athletic performance are frequently targeted for excessive

claims and irresponsible marketing. A number of Internet resources provide critical tools for patient education (Table 495-3). TABLE 495-3 Internet Resources on Complementary and Integrative Health Approaches The Cochrane Collaboration Complementary Medicine Reviews This website offers rigorous systematic reviews of mainstream and complementary health interventions using standardized methods. It includes

> 800 reviews of complementary health practices. Complete reviews require institutional or individual subscription, but summaries are available to the public. http://www.cochrane.org/evidence MedlinePlus All Herbs and Supplements, A–Z List MedlinePlus Complementary and Integrative Medicine MedlinePlus Dietary Supplements These National Library of Medicine (NLM) Web pages provide an A–Z database of science-based information on herbal and dietary supplements; basic facts about complementary and integrative health practices; and federal government sources on information about using natural products, dietary supplements, medicinal plants, and other complementary health modalities. http://www.nlm.nih.gov/medlineplus/druginfo/herb_All.html https://medlineplus.gov/complementaryandintegrativemedicine.html http://www.nlm.nih.gov/medlineplus/dietarysupplements.html National Institutes of Health National Center for Complementary and Integrative Health (NCCIH) This National Institutes of Health NCCIH website contains information for consumers and health care providers on many aspects of complementary and integrative health products and practices. Downloadable information sheets include short summaries of complementary health approaches, uses and risks of herbal therapies, and advice on wise use of dietary supplements. http://www.nccih.nih.gov Resources for Health Care Providers: http://www.nccih.nih.gov/health/providers NCCIH Clinical Digest e-Newsletter: http://www.nccih.nih.gov/health/providers/ digest Continuing medical education lectures: http://www.nccih.nih.gov/training/ videolectures Herbs at a Glance fact sheets: https://www.nccih.nih.gov/health/herbsataglance

Because many complementary health products and practices are used as self-care and because many patients research these interventions extensively on the Internet, directing patients to responsible websites can often be very helpful.

The scientific evidence regarding complementary therapies is frag mentary and incomplete. Nonetheless, in some areas, particularly pain management, it is increasingly possible to perform the kind of rigorous systematic reviews of complementary health therapies and practices that are the cornerstone of evidence-based medicine. A particularly valuable resource in this respect is the Cochrane Collaboration, which has performed >800 systematic reviews of complementary health prac tices. Practitioners will find this a valuable resource to answer patient questions. Practice guidelines, particularly for pain management, are also available from several professional organizations. Links to these resources are provided in Table 495-3. SUMMARY The frequent use of complementary and integrative health therapies and practices reflects an active interest among the public in improving health and well-being of the whole person. The current health care system

is fragmented, with diseases and comorbid conditions mostly treated separately, sometimes with drugs that interact with one another. An important step in whole person health care is considering health and disease not as separate states but as a bidirectional continuum and understanding how complementary and integrative therapies and prac tices, which are often multicomponent in nature, consider a patient’s long-term recovery and overall health. CHAPTER 495 Acknowledgment Dr. Josephine Briggs contributed to this chapter in prior editions and some material from prior edition chapters has been retained here. ■ ■FURTHER READING Black LI et al: Use of complementary health approaches among children Complementary and Integrative Therapies and Practices

aged 4–17 years in the United States: National Health Interview Survey, 2007-2012. National health statistics reports; no 78. Hyattsville, MD, National Center for Health Statistics, 2015. Eisenberg DM et al: Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. JAMA 280:1569, 1998. Gaston TE et al: “Natural” is not synonymous with “safe”: Toxicity of natural products alone and in combination with pharmaceutical agents. Regul Toxicol Pharmacol 113:104642, 2020. Ijaz N et al: Whole systems research methods in health care: A scoping review. J Altern Complement Med 25:S21, 2019. Nahin RL et al: Expenditures on complementary health approaches: United States, 2012. Natl Health Stat Rep 95:1, 2016. Nahin RL et al: Use of complementary health approaches overall and for pain management by US adults. JAMA 331:613, 2024. Paige NM et al: Association of spinal manipulative therapy with clini cal benefit and harm for acute low back pain: Systematic review and meta-analysis. JAMA 317:1451, 2017. Qaseem A et al: Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the American College of Physicians. Ann Intern Med 166:514, 2017. Skelly AC et al: Noninvasive nonpharmacological treatment for chronic pain: A systematic review up-date. Comparative Effective ness Review No. 227. AHRQ Publication No. 20-EHC009. Rockville, MD, Agency for Healthcare Research and Quality; April 2020. Vickers AJ et al: Acupuncture for chronic pain: Update of an indi vidual patient data meta-analysis. J Pain 19:455, 2018.

04 - 496 Placebo and Nocebo Effects

496 Placebo and Nocebo Effects

Kathryn T. Hall, Alia J. Crum

Placebo and Nocebo

Effects Placebos are sham versions of drugs, devices, or surgeries that lack the active compound, function, or procedure they are designed to simulate (Table 496-1). Administration of these “inactive” treatments can have significant therapeutic benefits called placebo effects, which accounts for their use as controls in clinical trials as comparators for active drugs, devices, or surgical interventions. Key drivers of placebo effects include the patient’s expectation and conscious or subconscious conditioning. Psychological studies demonstrated that expectations are shaped by factors intrinsic to the patient, including their past experi ences and core beliefs or mindsets, and extrinsic factors including environmental cues (e.g., white coat), clinical practice (e.g., physi cal examination), and information received about a treatment (e.g., expected benefits or side effects). Neuroimaging studies have identified consistent changes in the brain in response to placebo treatment that suggest that placebo effects work by integrating incoming information about extrinsic factors with prior experience and mindsets to update expectations of treatment benefit. When expectations are negative, TABLE 496-1 Glossary of Terms commonly used in Placebo Studies PART 20 Emerging Topics in Clinical Medicine TERM DEFINITION Additivity in clinical trials The assumption that placebo and drug treatment responses are additive is a fundamental assumption in clinical trials. However, there are notable exceptions to this assumption in pharmacogenomic and brain imaging studies where subsets of the population have been reported to have differential effects in placebo and drug treatment arms of a trial. Development and culture Our caregivers and social environment influence the psychological processes that underlie the placebo effect. These processes are continuously shaped throughout life by the ideas, institutions, and interactions that constitute the culture in which we live. Expectation A specific belief about the future based on a prediction of what is most likely to happen. Examples: “This drug will relieve my pain”; “I will experience side effects.” Gene-(drug/placebo) interaction Pharmacogenomic analysis has identified clinical trials in asthma, depression, pain, chronic fatigue, and cognitive function in which there are subpopulations based on genotype that have differential associations in the drug and placebo treatment arms. These differential effects often result in significant gene-(drug/placebo) interaction effects. Implicit learning The nonconscious acquisition of knowledge. Classical conditioning, a form of implicit learning, is implicated in certain instances of the placebo effect (e.g., implicit association of sleepiness with the administration of blue pills).

Mindset Core belief about a domain or category that orients an individual to a particular set of beliefs, associations, and expectations, and functions to guide attentional and motivational processes (e.g., “cancer is a catastrophe”; “symptoms are signs of efficacy”). Neurobiological mechanisms Dopamine, endogenous opioids, and endocannabinoids are three of the major neurotransmitter systems implicated in moderating the placebo effect. Placebo effects also work by activating biological properties of the body that facilitate healing, including homeostatic mechanisms and immune and inflammatory responses. These contribute to the natural history of a disease but can also be targets of placebo effects. Nocebo effect Sides effect or negative change in clinical outcome observed after exposure to negative information, interactions, or cues that can induce negative expectations. Open-label placebos (OLPs) OLPs are placebos administered to patients with their full knowledge that the treatment lacks the active pharmaceutical agent. OLP clinical trials have been conducted in irritable bowel syndrome, chronic back pain, allergic rhinitis, cancer-related fatigue, attention deficit hyperactivity disorder, major depression, and menopausal hot flushes. Meta-analysis of OLP trials found a significant overall effect. Patient-clinician relationship The patient-clinician relationship shapes the mindsets and expectations a patient holds about health, illness, and treatments, and affects the quality of care a patient receives. This relationship is influenced by the warmth and competence of the provider and is further shaped by characteristics like empathy and trust. Placebo Placebos are sham versions of drugs, devices, or surgeries that lack the active compound, function, or procedure they are designed to simulate. Placebos are often used in clinical trials as controls for placebo effects, natural history, regression to the mean, spontaneous remission, and Hawthorne effects (the tendency for people to change their behaviors when being observed). Placebo effect Positive change in clinical outcome observed after a placebo treatment; an exclusively attributed expectation mediated by psychological, neurological, or physiological placebo mechanisms. Placebo response Improvement observed among patients assigned to placebo treatment in a clinical trial. Placebome The genome-related products that modify placebo response. Several genes in neurotransmitter and other pathways have been implicated in modifying response to placebo treatment in clinical trials. The most well-studied of these is in the gene encoding catechol-Omethyltransferase (COMT). Social and observational learning Learning through direct observation of others undergoing treatment (i.e., other patients) and interactions with individuals who weild influence over the patient (i.e., physicians and nurses) both may powerfully drive placebo effects. Treatment characteristics The specific characteristics include factors like the shape, color, and labeling of the treatment; the method of administration; and the physical environment in which the treatment is administered.

they can result in negative outcomes, termed, “nocebo effects.” For many years, placebo effects were viewed as superfluous nuisance variables to be ignored, marginalized, and simply “controlled for” in clinical trials. With advances in psychology and neuroscience stud ies of placebo effects, the value of understanding their mechanism of action and harnessing these effects in clinical care and clinical trials has become increasingly apparent. INTRODUCTION TO PLACEBO EFFECTS:

A BRIEF HISTORY The word “placebo,” derived from the Latin placere, “to please,” first appeared in medical literature in clinical lectures of William Cul len, a leading physician in the eighteenth century. In 1792 he wrote, “I prescribed therefore in pure placebo, but I make it a rule even in employing placebos to give what would have a tendency to be of use to the patient.” Cullen was describing placebos being used to please, rather than treat; however, use of placebos was commonplace at that time, especially when effective therapies were exhausted or unavailable.

From the late eighteenth through the early twentieth centuries, sham treatments were also used to expose some unorthodox treatments as frauds—or, at least, as no better than placebos. In his highly publicized 1784 study of “animal magnetism,” a therapy developed by Austrian physician Anton Mesmer, Benjamin Franklin and a team of leading scientists in France simulated Mesmer’s elaborate rituals using fake practitioners and sham magnetism. This early placebo-controlled trial

demonstrated that the hugely popular and remarkably effective treat ment was no better than a placebo. Early trialists attributed the positive benefits of these therapies to the power of the “imagination,” reinforc ing the belief that placebo treatments were the product of wayward physicians, or “quacks,” who tricked gullible patients. After World War II, advances in metabolism, physiology, and clini cal pharmacology created a growing need for clinical trials to evaluate novel compounds for their ability to kill pathogens and alter disease processes. By the 1960s, the Declaration of Helsinki and KefauverHarris Amendments, introduced to protect patient safety by requiring informed consent, rendered use of placebos and the deception histori cally thought to promote placebo effects unethical, institutionalizing the transformation of placebos from salve to epistemic tool. Together, placebo controls, double-blinding, and randomization are considered the gold standard for acquiring the strongest evidence of efficacy or lack of efficacy for novel treatments. Notably, placebos are rarely used in trials of serious illnesses like cancer or when an effective treatment already exists. Then their use is limited to comparing novel treatments to standard of care plus placebo. In randomized placebo-controlled clinical trials, the effect of the drug is calculated by simply subtracting outcomes in the placebo treatment arm or placebo response from the drug response (Fig. 496-1A). Hence, in addition to controlling for placebo effects, placebos control for changes in the outcome of interest due to natural history (the tendency for a common cold to resolve on its own in 7–10 days), regression to the mean (a statistical phenomenon where extreme baseline measures tend to move toward the group mean), and the Hawthorne effect (the tendency for people to change behaviors when being observed). These variables, together with placebo effects, make up the placebo response. If the active treatment being tested significantly outperforms the pla cebo response, it is deemed efficacious and can progress through the U.S. Food and Drug Administration (FDA) approval process. However, if response to the drug is not statistically significantly greater than the placebo response, the active treatment is deemed lacking in efficacy. Clinical trials’ limited view of placebos obscures the reality that the effects associated with placebos are not, in practice, superfluous. Indeed, the effects of patient psychology (e.g., expectations, mindsets) and the social and cultural context (e.g., a clinician’s demeanor and drug label and advertising information) have a meaningful impact on health outcomes that warrant a more complete understanding. In addition, the effects of these factors are difficult to isolate. In fact, the total response to drug is the product of both the drug and the social and psychological context interacting with patient biology to bring about change (Fig. 496-1B). This perspective propels us into a new era of understanding placebo effects: not as treatment alternatives or as something to subtract, but as psychological, social, and biological mechanisms that can be considered an integral component of the overall treatment effect in medicine. By understanding placebo effects in this manner, we can optimize their benefits to improving drug discovery, maximizing existing treatments, and minimizing nocebo consequences to reduce harm. PLACEBO RESPONSE IN CLINICAL TRIALS Placebo effects, although often thought of in the context of a placebo pill, extend to many other treatment modalities, including sham surgeries, placebo acupuncture, and placebo diets. They have been documented in numerous conditions and diseases, including pain, depression, Parkinson’s disease, anxiety disorders,

cardiovascular dis orders, cancer-related fatigue, asthma, and gastrointestinal disorders. Not limited to patient-reported outcomes, placebo effects can affect objective physiologic outcomes, including blood pressure measure ments, immune biomarker levels, exercise endurance, and cognitive test scores. Recently, clinical trial sponsors have invested considerable resources in reducing the impact of placebo response by adjusting patient-level variables, such as reducing patient-clinician interactions and reducing patient expectations by providing neutral information about expected benefits and side effects. In conditions like Alzheimer's disease, trial ists have modeled placebo response over time and set the optimal

treatment duration length to the time period beyond the 8-week period when placebo response is maximal. The placebo run-in design attempts to eliminate the influence of placebo responders by assigning all enrolled blinded patients to a placebo at the beginning of a study; patients who respond to placebo are subsequently excluded. Other more complex models, such as the sequential parallel comparison design (SPCD), randomize patients to placebo or drug at a ratio of 2:1. After a brief treatment period, placebo nonresponders are rerandom ized to placebo or drug. Unlike placebo run-ins, SPCD uses all patients and, thus, should have greater power to find an effect. Still, despite these and other considerable investments, no approach reliably reduces the impact of placebo response on clinical trial failure.

MECHANISMS OF PLACEBO EFFECTS ■ ■PSYCHOLOGICAL MECHANISMS Expectations Expectations, or beliefs about the likelihood of future events, are thought to be the key driver of placebo effects. Expecta tions can be conscious; for example, many patients expect nonsteroidal anti-inflammatory drugs (NSAIDs) will relieve pain, melatonin will improve sleep, and beta blockers will reduce anxiety. Expectations can also be consciously or subconsciously conditioned. Repeated use of blue sleeping pills can induce sleepiness by taking a blue placebo pill. Multiple sclerosis patients who received the immunosuppressant cyclophosphamide paired with flavored syrup later displayed drugconsistent immune responses to the flavored syrup alone. Obser vational learning can also play a role in eliciting placebo effects by altering expectations. Watching another person experience pain relief in response to a particular treatment can lead the observer to expect to experience similar relief, even if the stimulus is a placebo in both cases. CHAPTER 496 Mindsets Mindsets are core beliefs about a broader domain or category, such as the meaning of side effects or the nature of a disease or treatment. Mindsets orient individuals to a set of associations, expectations, and goals. A mindset such as "cancer is a catastrophe," "statins are harmful," or "my body is permanently damaged" can shape a patient's experience of pain or other side effects. While mindsets and expectations are related, they are not identical. For example, a patient in pain may have the specific expectation that a treatment will relieve their pain. But they also could have broader mindsets in which those expectations are operating, such as "injections don't work," "my condi tion is hopeless," or "I am in good hands." Placebo and Nocebo Effects While specific expectations can be measured and assumed to influ ence placebo effects in studies, mindsets may be particularly powerful in the real-world practice of medicine where individual expectations do not exist in isolation. Mindsets may also be advantageous when con sidering how to harness placebo and minimize nocebo effects ethically. ■ ■SOCIAL AND CULTURAL MECHANISMS Language and Information Patients' implicit or explicit pre existing mindsets, shaped by the broader culture in which they were raised and/or reside, can be updated and informed through verbal instructions. In "open-hidden design" studies, medication adminis tered openly by a health care professional who informs the patient that they will experience benefit (e.g., "I'm going to administer a dose of

morphine, a powerful painkiller that will alleviate your pain") has a significantly greater analgesic effect compared to administering the same dose from a hidden pump without the patient's knowledge. Thus, even potent opioid analgesics lose as much as 30% of their efficacy if the patient is unaware that they received the treatment. Effects of openhidden paradigms are also seen with objective outcomes, such as heart rate. Information is conveyed not just by a clinician's words but also through information in the health care context more broadly, such as advertising and media related to drugs and health. Clinician Characteristics Beyond what the patient is told, trust in the source of information can also influence clinical outcomes. Socialpsychological research has shown that two qualities are key: patients' perceptions of competence, or whether a physician "gets it" (i.e.,

Treatment Response Drug Effect

Placebo Response

DRUG PLACEBO Clinical Trials Clinical Practice A PART 20 Emerging Topics in Clinical Medicine

Treatment Response

DRUG a/a PLACEBO a/a DRUG PLACEBO DRUG PLACEBO WHAT WE EXPECT TO SEE WHAT WE OFTEN SEE WHAT WE SEE WITH PHARMACOGENETICS B FIGURE 496-1 A. Additivity of drug and placebo response in a clinical trial. B. (1) What we expect to see: Expected outcomes from the classic view of clinical trials in which the effect of the drug exceeds placebo response. (2) What we often see: Typical results from a clinical trial in which there is no significant difference between the drug and placebo responses. (3) What we can see with pharmacogenomics: Pharmacogenomic analysis demonstrating differential effects of a genetic locus in the drug and placebo arms of a trial such that the drug effect and placebo response of one version of a variant (patients who are homozygous for the "A" allele, A/A) are opposite that of homozygotes of the alternate allele at that locus (patients who homozygous for the "a" allele, a/a). The average effects of outcomes of the two subpopulations cancel each other to give the results we often see. displays of efficiency, knowledge, and skill), and patients' perceptions of warmth, or whether a physician "gets me" (i.e., displays of personal engagement, connection, and care for the patient). Patients' assessments of clinician warmth and competence shape their treatment expectations and impact their mindsets about illness and, therefore, can influence placebo effects. In one study, an allergic

Drug Effects Pharmaceutical or physical properties of drug or intervention Placebo Effects PSYCHOLOGICAL MECHANISMS

• Expectations

• Mindsets

• Implicit Learning SOCIAL AND CULTURAL MECHANISMS

• Language and information

• Clinician characteristics

• Symbols

• Rituals

• Environmental Cues BIOLOGICAL MECHANISMS

• Genetic predispositions

• Neurobiological processes Other Effects OTHER EFFECTS THAT CAN INFLUENCE PLACEBO RESPONSE in TRIALS

• Statistical regression to mean

• Blinding and bias

• Informed consent and uncertainty

• Hawthorn Effects TREATMENT Placebo responder Drug responder (but drugnonresponder) (but placebononresponder) DRUG A/A PLACEBO A/A reaction was induced in participants via a histamine skin prick fol lowed by the administration of a placebo cream. The information about the cream was varied to create either positive expectations (“this cream will reduce your rash and irritation”) or negative expectations (“this cream may worsen your symptoms”). When the ensuing wheal was measured 10 min later, the difference in information alone produced

differences in the size of the allergic reaction. Interestingly, the effect of the information differed depending on perceived clinician warmth and competence. When the physician exhibited cues of both compe tence (e.g., wearing a white coat with a badge that read “Fellow at the Stanford Allergy Center”) and warmth (e.g., making eye contact with the patient), the effect of the spoken message was significantly enhanced. However, when social cues were changed to induce ques tions about the level of competence (e.g., badge read “Student Doctor”) and warmth (e.g., staring at a computer screen or making a personal connection), the information about the cream no longer mattered: both placebo and nocebo effects were minimized. There is no one right way to signal warmth and competence, but there are many ways to fail to convey these qualities and, therefore, lose patients’ trust. Indications of warmth and competence are important for building patients’ trust not only with their medical providers but also with the broader clinic, hospital, or health care system. Trust in the social context can magnify placebo effects or minimize nocebo effects; it can also have a direct effect on patient care, influencing a wide array of outcomes, such as metabolic complications, immune response, symptoms, and adherence to medication. Symbols, Rituals, and Environmental Cues Medical treat ment occurs within a rich context of environmental cues, symbols, and rituals. Many patients, for example, exhibit a transient (albeit substan tial) rise in blood pressure when in a medical setting, a phenomenon known as the “white coat” hypertension. Seemingly inconsequential characteristics of a drug, such as color, drug brand name, and cost, have been found to impact treatment efficacy. Patients tend to perceive capsules as stronger and more effective than tablets and tend to have a reduced response to placebos referred to as discounted or generic. In a within-subjects, repeated-measures study to examine drug labeling in migraine, patients were given either placebo or 10-mg rizatriptan labeled to create three information conditions ranging from negative (“placebo”), to neutral (“Maxalt or placebo”) to

positive (“Maxalt”). While patients had significantly greater relief from Maxalt labeled as Maxalt compared to placebo labeled as placebo, there was no difference in the effect of Maxalt labeled as placebo compared to placebo labeled as Maxalt. These findings demonstrate the ability of labeling and brand names to influence the effect of a drug. Interestingly, drug companies have increased their use of the letters X and Z in drug names as stud ies show these visually distinct letters have fewer negative associations with other medications. Furthermore, marketing research has found Z is associated with efficacy, whereas T and S are associated with greater tolerability. ■ ■BIOLOGICAL MECHANISMS Pharmacologic evidence from the 1970s showing that naloxone, an opioid antagonist, could abrogate placebo effects in the experience of pain after molar tooth extraction laid the groundwork for demonstrat ing that psychological forces could affect patient physiology. Early neuroimaging studies revealed the release of endogenous opioids and dopamine signaling proportionate to the expectation and perception of how well a given placebo intervention worked. Using models of placebo analgesia, neuroscientists imaged the brains of healthy subjects exposed to various forms of thermal or mechanical pain-producing stimuli in the presence or absence of placebo treatments (e.g., inert creams or inactive transcutaneous electrical nerve stimulation devices) to induce placebo effects using conditioning or suggestion. These stud ies revealed activation of regions in the spinal cord and descending pain-control regions in the brainstem and reduced signaling in the spi nothalamic tract. Later meta-analyses of 20 of these studies confirmed reduced signaling proportionate to reported placebo analgesia in painrelated activity in the thalamus, insula, and habenula, while increased signaling was observed in frontal-parietal regions. Today, placebo treatments are hypothesized to influence brain sys tems involved in “meaning-making” by constructing internal models that guide our understanding of incoming signals and their source, as well as implications for anticipatory events. These internal models provide predictive signals that are incorporated with incoming sensory

signals to produce the sensations and symptoms experienced. In turn, these models inform our perceptions (mindsets) and shape our reac tions, amplifying or attenuating perceptual and affective circuits. Thus, contextual information around placebo treatments, including the sug gestion of benefit, the visual and behavioral cues provided via the ritual of treatment, and prior experiences of benefit, can modify the neural construction of the experience and, in turn, downstream physiologic consequences in the nervous, immune, endocrine, and cardiovascular systems.

THE CHALLENGE OF IDENTIFYING “PLACEBO RESPONDERS” In drug development, identifying and excluding placebo responders could lead to more precise, and potentially smaller and less expensive, clinical trials. Predicting which patients are likely to respond to pla cebos could allow clinicians to optimize patient interactions and even support gradual replacement of drugs with side effects with placebo by dosage titration. Research into psychological predictors of placebo responders found several personality traits and constructs to be associated with placebo responders, including optimism, habitual desire for control, fun and sensation seeking, neuroticism, self-efficacy, and internal locus of control. Functional magnetic resonance imaging (fMRI) has also been used to create brain-signaling profiles that are predictive of placebo responders. In a study of patients with chronic osteoarthritis pain, right midfrontal gyrus connectivity effectively identified placebo pill responders. Interestingly, in some subjects, the active drug in this study, duloxetine, appeared to enhance the placebo response, but in others, duloxetine reduced it. This finding suggests that while drug and placebo were additive for some patients, the interaction with placebo

response diminished the drug effect for others. CHAPTER 496 The observation that placebo effects are influenced by opioid and dopaminergic signaling suggested that genetic variation in the synthe sis, function, or metabolism of these neurotransmitters might influence the magnitude of placebo effects. This observation gave rise to the term “placebome” to describe the group of genome-related products that potentially influence individual response to placebo treatments. Mem bers of the placebome were identified in candidate and genome-wide association studies (GWAS) of the placebo control arms of clinical tri als. For example, there is evidence of differential effects associated with the genes COMT and MAO-A, which encode enzymes that metabolize dopamine; OPRM1, which encodes the opioid receptor; and TPH2 and HTR2A, which encode proteins involved in serotonin signaling. Of these genes, COMT has the strongest evidence for association with pla cebo response in clinical trials of irritable bowel syndrome (IBS) and pain. To date, there are 29 genes associated with response to placebo in the GWAS catalog. Placebo and Nocebo Effects Although numerous psychological, neuroimaging, and genetic profiles of placebo responders have been proposed, these biomarkers were mostly derived from small studies and have not yielded consistent results in prospective studies. This is in large part due to the broad het erogeneity in variables intrinsic to patients, including their conditions and disease severity and duration, and extrinsic study variables, includ ing treatment duration, inclusion criteria, study location, number of study visits, outcome measures, and information about the study drug (e.g., possible side effects). ■ ■ADDITIVITY IN CLINICAL TRIALS The study of placebo responders has also led to important evidence of nonadditivity in clinical trials. Additivity between drug and pla cebo outcomes has been a universal and fundamental, but unproven, assumption in clinical trials. Based on additivity, the drug effect is determined by subtracting the outcome in the placebo arm from the outcome in the drug treatment arm (Fig. 496-1A). As reported in the study investigating brain connectivity of patients with depression described above, in some patients, duloxetine enhanced brain connec tivity seen in the brain region associated with placebo response. How ever, in other patients, this connectivity was reduced with duloxetine. Similarly, in clinical trials of chronic pain, chronic fatigue syndrome,

cardiovascular disease, and asthma, significant genetic associations observed in placebo arms were null or found to be in the opposite direction in drug treatment arms. These unexpected gene-by-(drug/ placebo) interaction effects suggest that, in some clinical trials, there are subsets of patients for whom the drug and placebo response is not additive (Fig. 496-1B). Thus, the potential for differential outcomes in the drug and placebo arms could confound outcomes in clinical trials, warranting further investigation.

NOCEBO EFFECTS In part, because informed consent in clinical trials requires disclosure of all potential drug side effects, the side effects reported by patients randomized to placebo are often the same as those expected with ran domization to drug. When this phenomenon was first documented in 1961, the word nocebo, coined from the Latin nocere, “to harm,” was used to describe production of negative effects from negative verbal suggestions, contextual cues, or associative learning. Although the term nocebo was originally defined as an adverse effect from an inert treatment, nocebo effects, like placebo effects, are the product of underlying aspects of the patients’ psychology (in this case, negative expectations and mindsets) and the social context. In clinical trials, on average, 25% of participants randomized to placebo report side effects, and some studies (such as those of statins) show that the rates of side effects do not significantly differ between the active drug and placebo. Because they did not receive the active drug, we can

assume the side effects arose, in part, due to expectation and not due to any active ingredients in the treatment. Nocebo effects are not limited to clinical trials. While statins are effec tive cholesterol-lowering agents, the belief that they cause muscle pain is widespread, and treatment is often discontinued because of this reported side effect. This phenomenon has been extensively investigated. In one study, a “within-subjects” design was used in which each patient served as their own control. In this study, patients who reported side effects were blinded and randomized to take a placebo, statin, or no treatment on a monthly basis over a 1-year period. At the end of the study, there was no discernable difference between symptoms reported on placebo versus statin, and 50% of the patients in the trial were able to reinitiate statin therapy successfully. Negative expectations surrounding technology (e.g., wi-fi or cell phone signals), environmental agents (e.g., infrasound generated by wind turbines), or food (e.g., gluten) can also enhance the likelihood of negative symptoms related to their presence. PART 20 Emerging Topics in Clinical Medicine Ethically, it is hard to test nocebo effects deliberately, but random ized studies in laboratory contexts show that people who are told to expect side effects are more likely to experience those side effects than those who are not told to expect side effects. Expectations and mindsets can deepen negative effects through physiologic activation or by heightening awareness of symptoms that may have already been present, resulting in misattribution of their cause. Nocebo responses can also occur because of conditioned learning. For example, patients receiving chemotherapy can develop nausea when they see or smell a stimulus associated with their treatment, such as the treatment room or even a staff member. Key drivers of placebo and nocebo effects overlap with factors that create barriers to quality clinical care for black patients and other patients of color. Poor communication, perceived discrimination, and medical mistrust are all factors demonstrated to reduce the quality of care in racially discordant dyads. Prior experiences of discrimination in the health care setting may result in expectations of discrimination and suboptimal clinician communication, enhancing nocebo and reducing placebo effects during treatment encounters. Consequently, the pres ence of nocebogenic and absence of placebogenic influences associ ated with racially discordant dyads has the potential to generate and exacerbate racial and ethnic inequities in clinical outcomes and care. ETHICALLY AND DELIBERATELY HARNESSING PLACEBO EFFECTS If placebo effects are understood as an integral component of the overall treatment effect, mediated by neurobiological processes and social and environmental factors, they can be personalized and maximized in the

practice of medicine. Administering placebos without full knowledge of the patient is no longer acceptable for important ethical reasons. Moreover, administering “impure placebos”—pharmacologically active treatments that are prescribed at too low a dose to be effective or are known to be ineffective for the condition being treated—is reportedly common practice but still controversial and ethically problematic. The use of impure placebos varies by country. In the United States, a survey of 1200 internists and rheumatologists indicated that 62% of participants believed the practice of utilizing impure placebos—over-the-counter analgesics and vitamins—was ethically permissible. Notably, <5% reported using saline or sugar pills. The numbers are higher in Canada and the United Kingdom, where 80 and 77% respectively, of physicians surveyed reported prescribing impure placebos or treatments without proven or expected benefit. In Denmark, 86% of internists, 54% of hospitalists, and 41% of specialists in private practice report that they used placebo interventions at least one time within the previous year. Finally, the German Medical Association, after assessing placebos in medicine, published a report in 2011 acknowledging the complexity of the strong effects that placebos can have,

supporting their limited use when no other therapies were available. In addition to using impure placebos, there are several other alternatives that deliberately leverage benefits of placebo effects. ■ ■OPEN-LABEL PLACEBOS One way that researchers have addressed the ethical and legal limita tions on placebos is to simply tell patients the facts about placebos. In honest or open-label placebo (OLP) studies, patients are fully informed about the absence of the active agent in the placebo, but are also told that placebo treatments can sometimes result in clinical improvement. The patients are also informed that trying a placebo might yield some benefit; even if they do not believe this to be the case, they could con sider suspending their belief. To date, the findings on the effects of OLP are promising. Improvements have been observed in IBS, cancerrelated pain and fatigue, depression, posttraumatic memory intrusions, allergic rhinitis, attention deficits, and hyperactivity; however, OLP has yielded no benefit in wound healing and did not enhance the cognitive abilities of healthy volunteers. ■ ■DELIBERATELY LEVERAGING PATIENT EXPECTATIONS AND MINDSETS When understood as being driven by the psychological and social context, placebo effects can be evoked without the use of sham pills or procedures by deliberately shaping patient expectations and mindsets. At their best, doctors and patients alike already harness the forces behind placebo effects through interactions, branding, and language that inspire patients' trust, as well as useful mindsets and expectations about their condition. Once aware of this fact, health care practitioners can work ethically to leverage these forces to improve health outcomes. In the PsyHEART trial, 124 patients undergoing coronary artery bypass surgery were randomized to standard care, supportive therapy, or an expectation of manipulation in which they were encouraged to develop clear expectations about how their life would improve after surgery (i.e., what activities they would be able to perform). Six months after surgery, patients who were randomized to the expectation manipulation showed significantly greater improvements in quality of life and reductions in disability. In the EMBRACE study, patients diag nosed with cancer were exposed to documentary-style films featuring experts in psychology and oncology and to cancer survivors who spoke about how their mindsets changed throughout and after their treatment, as well as challenges they faced along the way. Participants exposed to the films increased their health-related quality of life, such as their emotional well-being, physical health, and general functioning, by 10%, as measured by changes in industry-standard scales, compared to patients receiving treatment as usual. While these interventions were delivered directly to patients, trainings to help physicians and care teams more deliberately and effectively shape patients' expectations and mindsets in the context of their care are being developed, evalu ated, and disseminated.

05 - 497 The Role of Epigenetics in Disease and Treatment

497 The Role of Epigenetics in Disease and Treatment

■ ■ETHICAL CONSIDERATIONS In all of these applications, ethical considerations are of utmost impor tance; it is never acceptable to deceive or provide false information to the patient. Within these bounds, however, there is much we can do to improve patients' mindsets and expectations. Consider the nocebo effects resulting from informing patients about side effects. While it is not ethical to withhold this information from patients, providers could either provide more realistic expectations about the likelihood of side effects and set more adaptive mindsets about their meaning. In one study of children undergoing oral immunotherapy treatment (OIT) for peanut allergies, half were randomized to receive a typical warning message: side effects are negative outcomes, unrelated to treatment efficacy, that need to be managed and endured. The other half were given messages aimed to instill the mindset that some mild symptoms are often a sign that the treatment is working and signal desensitization. Compared with families informed that symptoms are negative side effects, families informed that "symptoms are positive signs of treatment efficacy" experienced significantly less anxiety, fewer symptoms during the highest doses, and improved levels of IgG4, an immune marker of allergic tolerance. Similar effects of this messaging have proven to reduce anxiety and side effects for those receiving the COVID-19 vaccine. THE FUTURE OF PLACEBO EFFECTS We are entering a new era of understanding about placebo effects, one in which they are not viewed as treatment alternatives or as something to subtract, but as psychological, social, and biological mechanisms that can be considered an integral component of the overall treatment effect in medicine. Work in this field is proliferating, and translation of the findings to clinical trials and clinical care is important for optimiz ing placebo effects to improve existing treatments while minimizing nocebo effects to reduce harm. ■ ■FURTHER READING Colloca L et al: Placebo Effects through the Lens of Translational Research. New York, NY: Oxford University Press, 2023. Evers AW et al: Implications of placebo and nocebo effects for clinical practice: Expert consensus. Psychother Psychosom 87:204,

)18.

Hall KT et al: Systems pharmacogenomics: Gene, disease, drug and placebo interactions: A case study in COMT. Pharmacogenomics 20:529, 2019. Hall KT: Placebos. MIT Press Essential Knowledge Series. Cambridge, Massachusetts: The MIT Press, 2022. Howe L et al: Changing patient mindsets about non-life-threatening symptoms during oral immunotherapy: A randomized clinical trial.

J Allergy Clin Immunol Pract 7:1550, 2019. Petrie KJ, Rief W: Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. Annu Rev Psychol 70:599, 2019. Rief W et al: Preoperative optimization of patient expectations improves long-term outcome in heart surgery patients: Results of the randomized controlled PSY-HEART trial. BMC Med 15:4,

)17.
Zion SR, Crum AJ: Mindsets matter: A new framework for harnessing the placebo effect in modern medicine. Int Rev Neurobiol 138:137, 2018. Zion SR et al: Changing cancer mindsets: A randomized controlled feasibility and efficacy trial. Psychooncology 32:1433, 2023. Zunhammer M et al: Meta-analysis of neural systems underlying pla cebo analgesia from individual participant fMRIData. Nat Commun 12:1391, 2021.

Brian C. Capell, Shelley L. Berger

The Role of Epigenetics

in Disease and Treatment The term epigenetics was coined by Conrad Waddington in 1942, as he sought to explain how changes in phenotype could occur throughout development independent of any changes to genotype. Appending the prefix epi- (Greek, meaning “over, outside of, around”) to genet ics aptly describes the numerous mechanisms by which gene expres sion and phenotypes are influenced—and sometimes even inherited through cell division—independent of any changes to the underlying DNA sequence. Today, epigenetics occupies one of the most exciting topics in biology and medicine, offering profound opportunities for discovery, as well as promise for the development of new therapies for disease. Interdisciplinary by nature, the field crosses virtually all areas of science and medicine: chemistry and genetics, development and dif ferentiation, immunology, cancer, aging, and neuroscience. The continuous introduction of ever more powerful technologies for interrogating the epigenome has led epigenetics to become one of the most innovative fields within the biomedical sciences. Given the vast expanse of the topic and limitations of space, in this chapter, we provide a broad but brief overview of the field and then highlight key areas across the landscape of biomedicine where epigenetics has been revealed to play critical roles in physiology and disease, and impor tantly, where epigenetics-based therapies have demonstrated success in clinical medicine. CHAPTER 497 ■ ■THE BIOCHEMICAL BASES OF EPIGENETICS Fundamental to epigenetic regulation is the intricate organization into chromatin of each cell’s genome (Chap. 479). The fundamental unit of the packaging into chromatin is the nucleosome, consisting of 147 base pairs of DNA wrapped around an octamer of 8 histone proteins (two copies of each of the four core histone proteins: H2A, H2B, H3, and H4), and nucleosome assembly into a regular repeating spaced array along the DNA polymer. The presence of nucleosomes and level of compaction of this basic chromatin array determine the accessibility of the DNA strand to transcription factors, to DNA repair machinery, and to other DNA-binding entities. Thus, compaction has a profound influence on gene expression levels and on local DNA

mutation rates. Open regions of chromatin (euchromatin) tend to be transcriptionally active, whereas compacted chromatin (heterochromatin) tends to be transcriptionally repressed. Higher order three-dimensional chromatin architecture such as folding and looping further contribute to epigen etic gene regulation and cellular phenotypes. The Role of Epigenetics in Disease and Treatment Histones include the four core histones, which are the most abun dant and most frequently found throughout the genome, and the vari ant histones of H2A, H2B, and H3. The individual protein structures of both core and variant histones include amino- and carboxyl-terminal "tails," which are extended and unstructured, and highly conserved globular domains. The x-ray crystal structure of the nucleosome par ticle has illuminated the dynamic alterations of chromatin by an aston ishing range of regulatory mechanisms, summarized below. The three main processes that regulate chromatin compaction, and thus access to the DNA template, include direct methylation modifi cations (and oxidized derivatives of methylation) of the DNA strand itself, posttranslational modifications of histones, and remodeling of nucleosomes to alter their location and composition with variant histones (Fig. 497-1). The major modification of DNA is cytosine methylation of CpG dinucleotides (5-mC), associated with gene repres sion and catalyzed by the DNMT1, DNMT3A, and DNMT3B enzymes. DNMT3A and 3B catalyze the addition of methyl groups on unmeth ylated DNA de novo at CpG dinucleotides that are typically located throughout transcribed genes and in intergenic regions, but lacking at promoters, while DNMT1 is critical for the maintenance of the meth ylation state after DNA replication and after transcription during the

Tonsils Thymus Bone marrow Lymph nodes Spleen Appendix IMMUNE SYSTEM Chromosome PART 20 Emerging Topics in Clinical Medicine DEVELOPMENT AGING METABOLISM CANCER FIGURE 497-1 Epigenetic pathways influence multiple physiologic and disease pathways. As depicted in the center of the illustration, epigenetics refers to the chemical modifications of DNA and histones, which influence chromatin structure, gene expression, and susceptibility to mutations. These molecular pathways, in turn, play important roles in development, cancer, metabolism, aging, neural function, and behavior, and in the immune system. ETC, electron transport chain; TCA, tricarboxylic acid. S phase of the cell cycle. To further alter and to remove methylation, the TET enzymes (TET1–3) progressively oxidize 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), to 5-formylcytosine (5-fC), and to 5-carboxylcytosine (5-caC), which are unable to be recognized by DNMT1 but can be removed by additional enzymes. Hence, these are mechanisms to passively lose 5-mC following DNA replication or to actively remove 5-mC, both potentially returning to unmethylated cytosine. Histone posttranslational modifications (hPTMs) are rich sources of diverse signaling to, and marking of, the chromatin template, includ ing at least 60 different covalent chemical modifications on the histone N- and C-terminal tails and within the globular domains. The hPTMs are added (written) and removed (erased) by enzymes and also serve as sequence- and PTM-specific binding surfaces for effector proteins and complexes (readers) to carry out a wide range of downstream actions including transcription, replication, DNA repair, and recombination.

BRAIN AND BEHAVIOR DNA methylation Histone methylation Histone acetylation DNA Nucleosome Glucose TCA ATP ETC One key point is that the staggering numbers of writers, erasers, and readers provide unlimited potential for diagnostic and therapeutic pharmacologic discovery. Throughout this chapter, we focus on histone acetylation and methylation, the most abundant and the most well-studied hPTMs (Fig. 497-1), although a wealth of additional modifications, such as serine/threonine/tyrosine phosphorylation, lysine ubiquitination, lysine SUMOylation, and lysine

ADP-ribosylation, among others, play important roles in transcriptional and chromatin regulation. For instance, histone phosphorylation targets histone H2A at Ser139 (γH2A.X), which marks DNA double-strand breaks immediately fol lowing DNA damage and is critical for the recruitment of the DNA repair machinery. Histone mono-ubiquitination functions similarly to other hPTMs, in signaling and marking the chromatin template, in particular serving to mark the initiation region or elongation of transcribed genes for future rounds of transcription, whereas histone

SUMOylation plays a role in transcriptional repression. Polyubiq uitination serves to tag proteins for degradation by the proteasome, and dysfunction in this system may play a role in the pathogenesis of neurodegenerative diseases, including Alzheimer's, Parkinson's, and Huntington's. ADP-ribosylation involves a class of enzymes, the polyADP-ribose polymerases (PARPs), which transfer ADP-ribose units from NAD+ to a variety of nuclear proteins. This PARylation alters the chromatin environment through the recruitment and modification of chromatin-associated proteins. In general, future studies of the profuse types and functions of hPTMs will enhance our understanding of these chromatin-based mechanisms and processes and will illuminate new opportunities and targets for therapies. In contrast, there is extensive understanding of histone lysine acetyl transferases (KATs) and methyltransferases (KMTs). KATs, previously known as HATs, were among the first identified histone modification enzymes. They attach acetyl groups on the lysine residues of histone tails and other proteins, resulting in both a novel side chain (acetyllysine) and an increase in negative charge (from positive charged lysine to neutral acetyl-lysine). This alteration results in loosening of chromatin structure to become more permissive to the binding of tran scription factors, and acetylation also creates a novel binding surface for the association of reader proteins. Acetylation on core histones, such as lysine 9 on histone H3 (H3K9ac) or lysine 27 (H3K27ac), is typically associated with transcriptional activation. Acetylation is very dynamic and can be rapidly removed by histone deacetylases (HDACs), of which there are multiple classes, including HDACs and sirtuins (NAD-dependent deacetylases), acting to return the lysine to unmodified ground state. Methylation of histone tails by KMTs provides more nuanced regulation, in that particular methylated lysines are associated with transcriptional activation (e.g., H3K4me3, H3K36me3, H3K79me3), transcriptional repression (e.g., H3K27me3), or DNA repeat and cen tromeric silencing (e.g., H3K9me3). The output is strictly determined by effector protein binding, as methylation of lysine does not alter side chain electrostatic charge. Lysine methylation is also a more stable chemical modification than is acetylation and turns over more slowly. Lysine demethylases have been identified for several of the specific methylated sites (H3K4, H3K9, H3K36, H3K27, H3K79). In addition to their impacts upon local chromatin structure through electrostatic alterations and through recruitment of reader effector proteins, some histone modifications can influence other epigenetic processes. For example, H3K36me3 is involved in a variety of tran scriptional processes including elongation and splicing. However, through its recruitment and interaction with other methyltransferases, such as DNMT3B and METTL14, H3K36me3 impacts both DNA and RNA methylation, respectively. Frequently coordinating with histone modification enzymes are nucleosome remodeling enzymes, which use the energy derived from the hydrolysis of ATP to reposition and remove nucleosomes along the DNA template and to exchange core histones and variant histones (including variants that are located at the transcriptional initiation sites [H2AZ] and over the transcribed genes [H3.3]). The nucleosome remodeling complexes can activate or repress transcription. The SWI/ SNF family creates nucleosome-free regions for transcriptional activa tion, the ISWI family evenly spaces nucleosomes to repress transcrip tion, and the INO80 family exchanges H2A with H2AZ at transcription start

sites to poise transcriptional activation. Other remodeling com plexes play key roles in the DNA damage response and apoptosis, among additional genomic processes. As alluded to above, RNA can also be methylated, and “RNA epi genetics” is now an emerging area of gene regulation beyond the direct methylation of DNA and hPTMs. Methylation of RNA, such as mes senger RNAs (mRNAs), has been known to exist for over half a century. However, in the last decade, the discovery of enzymes that perform reversible methylation of RNAs led to an explosion of this new field, called epitranscriptomics. Indeed, RNA methylation leads to mRNA degradation or facilitates translation. However, mRNA methylation itself occurs co-transcriptionally. Notably, the writer methyltransferase enzymes (METTL3, METTL14) and the demethylases (ALKBH5, FTO)

have important roles in a variety of disease pathologies, and drugs target ing their clinical activities are currently in human clinical trials.

Because multiple enzymes redundantly and synergistically write, erase, and recognize these modifications on DNA, RNA, and histones, there is great complexity and the potential for fine-tuning of gene regulation. While extensive knowledge gaps remain to fully explicate these mechanisms of gene regulation, epigenetics has become a fully established discipline within biomedical research. In the coming years, it is likely that the basic understanding of these processes will be fur ther harnessed for further betterment of human health. ■ ■TOOLS FOR THE STUDY OF EPIGENETICS Central to the rapid pace of epigenetic discovery has been the continual development of new cutting-edge epigenetic technologies. Chromatin immunoprecipitation (ChIP), developed over three decades ago, has been a mainstay across epigenetics and molecular biology research more broadly (Fig. 497-2). ChIP involves using formaldehyde to cross link proteins to DNA and then fragmenting the DNA and reversing of the crosslinks in order to analyze the DNA. The linking of ChIP to next-generation sequencing (ChIP-seq) provided a major leap forward in allowing researchers to probe the entire genome-wide landscape of histone modifications and DNA-binding transcription factors and chromatin-modifying enzymes. This has led to fundamental discov eries regarding the role of the epigenome in the regulation of gene expression and cellular phenotypes in development and disease. More recent refinements in these methods have expanded the appli cability of these methods. Studying chromatin accessibility has become possible through the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Fig. 497-2). Through ATAC-seq, a Tn5 trans posase can be utilized to insert sequencing adapters into open regions of chromatin, allowing for the identification of DNA regulatory elements such as promoters and enhancers even at the single-cell level. Building upon both ChIP-seq and ATAC-seq, the tethering of an antibody of interest to micrococcal nuclease (MNase) allows for the cleavage of the DNA on either side of the target, sidestepping the need for any formal dehyde fixation step, significantly scaling down the signal-to-noise ratio, and reducing the number of cells and DNA required in comparison to standard ChIP-seq. This method, referred to as CUT&RUN (cleavage under targets and release using nuclease), offers the ability to obtain his tone and chromatin-binding information in systems and models where cell numbers were previously rate limiting. A further modification of the CUT&RUN protocol replaces the MNase with a Tn5 transposase fused to sequencing adapters (CUT&Tag), offering the ability to profile histone modifications at the single-cell level. CHAPTER 497 The Role of Epigenetics in Disease and Treatment While ATAC- and ChIP-seq and their derivatives have provided tremendous insights into how chromatin accessibility and histone modi fications play a role in gene regulation, they did not provide information on how the physical organization and folding of the genome might con tribute to gene expression. This was only able to begin to be

understood by techniques that could elucidate the three-dimensional architecture and structure of chromatin. Here, techniques such as HiC (Hi-C is a high-throughput form of 3C (chromosome conformation capture) technology to study 3D genome organization) and Hi-ChIP (Hi-ChIP combines Hi-C with ChIP-sequencing to study the relationship of DNA-binding proteins to 3D genome organization) have emerged to reveal nuclear architecture and how it can either inhibit or facilitate gene expression. Collectively, these studies have revealed a model whereby enhancer state drives gene regulation. Once enhancer-promoter loops have formed, these topologically associated domains (TADs) are rein forced and can ultimately help to constrain motion of the genome and in turn increase the likelihood of further promoter-enhancer connections forming to facilitate transcription. More recently, the latest frontier in biomedical research is spatial technologies that allow the capture of molecular data at subcellular resolution within their native tissue context. While techniques such as spatial CUT&Tag are still in development, as they continue to advance in resolution, throughput, and accessibility, they are certain to offer unprecedented insights into how tissue and disease pathology corre lates with alterations in the transcriptome, epigenome, and proteome.

ChIP–seq ATAC-seq Epigenome editing UV Tn5 transposase Peaks (kb) Sequencing peaks corresponding to open chromatin Purified DNA Adapter Data collection Sequencing library PART 20 Emerging Topics in Clinical Medicine Reads: …GTTCCTTCAGCATTTGCAGCGT… Reads: Reference Genome Peak identification NGS Sequencing Motif 1 Motif 2 FIGURE 497-2 Core experimental techniques for the study of epigenetics. The explosion of interest and research in the past few decades has been fueled by fundamental advances in the experimental approaches and ability to profile the epigenome. Chromatin immunoprecipitation–next-generation sequencing (ChIP-seq) allows for the ability determine the genome-wide binding of a histone modification or DNA-binding protein of interest. In contrast, assay for transposase-accessible chromatin using sequencing (ATAC-seq) provides a method for determining chromatin accessibility genome-wide even down to the single-cell level. More recently, the development of CRISPR-based epigenome-editing technologies has offered a way to directly deposit histone modification in order to activate or repress specific genes. NGS, next-generation sequencing. Finally, another major technological breakthrough, and one with tremendous therapeutic potential, is development of CRISPR-based epigenome editing approaches (Fig. 497-2). By fusing a nucleasedeactivated Cas protein to an epigenetic modifying enzyme, one can use guide RNAs to precisely target gene regulatory effectors to turn on and turn off specific genes by changing the gene’s histone acetylation or methylation levels. For example, using CRISPR to guide the mRNA of an epigenetic repressor to the oncogene MYC is one strategy currently being tested for cancer treatment. These advances promise to not only elucidate new knowledge regarding principles of gene regulation but also to offer new therapeutic opportunities for disease. ■ ■EPIGENETICS IN DEVELOPMENT

AND DIFFERENTIATION Epigenetic processes are critical to organismal development and to cellular differentiation and reprogramming of cell fate (Fig. 497-1). Transcription factors establish the epigenomic landscape that enables and stabilizes cell-type-specific gene expression while simultane ously ensuring stable repression of alternative cell fates. This results in chromatin profiles that display remarkable cell-type specificity in

dCas9 Effector domain Epigenetic modifications Repressed locus dCas9 Activated locus Epigenetic editing differentiated cells, particularly at the key regulatory nodes of gene enhancers, which are

gene-distal DNA elements that control transcrip tion. In fact, epigenome profiling of the chromatin landscape in tumors of unknown cell origin can provide a better index of the origin tissue than does DNA sequencing of gene mutations within the tumor. The cell-type-specific epigenetic program is first derived from the template of embryonic stem cells, where numerous genes required for differentiation exist in a “bivalent” state, marked by both the activat ing histone modification, H3K4me3, and the repressive modification, H3K27me3. Due to this unstable epigenetic state, the genes are “poised” for activation or for repression, depending on their subsequent cell fate. Critical genes directing toward a specific cell fate will be turned on, with maintained H3K4me3 and erased H3K27me3, whereas genes leading toward alternative fates will be repressed, with maintained H3K27me3 and removed H3K4me3. Once differentiated, an epigenetic barrier will prevent the cells from returning to the stem cell state. For example, con stitutive heterochromatin in the form of H3K9me3 can serve as a barrier to cellular reprogramming when attempting to create induced pluripo tent stem cells, and inhibiting the enzymes that catalyze H3K9me3, such as SUV39H1, can enhance reprogramming efficiency.

DNA methylation contributes to the specification of cell fate and to other developmental pathways. DNA methylation alterations are involved in critical processes ranging from sex chromosome dosage compensation to coordinating expression of imprinted genes. Disrup tion of this latter process can lead to imprinting disorders including Prader-Willi syndrome, Angelman syndrome, and Beckwith-Wie demann syndrome. Recent discoveries have served to highlight the tremendous amount of interplay between epigenetic modifications and, in particular, between DNA methylation and various histone modifications. Beyond embryonic development, epigenetics can provide the neces sary coordination and balance between adult stem cell self-renewal compared to cell differentiation. This epigenetic control is critical, as impaired self-renewal can lead to stem cell exhaustion and premature aging, while excessive self-renewal may promote cancer. Key epigenetic regulators tend to play conserved roles across diverse tissue types. For instance, BMI1, a component of the polycomb repressive complex 1 (PRC1), is required for stem cell proliferation and self-renewal, and its ablation leads to stem cell depletion in hematopoietic, epidermal, muscle, intestinal, and mammary stem cells. Similarly, the DNA meth yltransferase DNMT1 is required for stem cell self-renewal in hemato poietic, epidermal, and mammary stem cells. HDACs 1 and 2 possess some overlapping functions and are required for normal epidermal differentiation. Likewise, a loss of these HDAC enzymes in hematopoi etic stem cells can lead to failure of differentiation and severe anemia. In a similar fashion, inhibition or loss of histone lysine demethylase 1 (LSD1), a repressor of transcription, is known to promote differ entiation across multiple cellular contexts. These factors represent repressive chromatin regulation, leading to the general concept that restraining specific transcription pathways related to differentiation is crucial to maintaining undifferentiated self-renewing stem cell pools. The epigenetic regulation of the tumor suppressor p16 (CDKN2A) locus during differentiation provides a prime example of this finely tuned system. For example, as mentioned above, DNMT1 is necessary for self-renewal in human epidermal stem cells. Levels of DNMT1 are high in the basal undifferentiated layer of the epidermis, decreasing progressively with epidermal stratification, leading to de-repression of the tumor suppressors p16 and p15, thereby promoting cell cycle arrest and full differentiation. BMI1 displays a similar phenotype in both hematopoietic and epidermal stem cells, repressing key genes that promote differentiation, such as p16 and p19ARF. Consistently, a loss of BMI1 leads to premature differentiation and defective self-renewal. In addition to the repression provided by DNMT1 and BMI1, the p16 locus is highly decorated with the repressive H3K27me3 catalyzed by EZH2 in

epidermal stem cells. Then, during epidermal differentiation, H3K27me3 is removed by the KDM6B (JMJD3) histone demethyl ase. Loss of this control over programmed p16 expression occurs in epithelial cancers, such as squamous cell carcinoma (SCC), where EZH2 is overexpressed and KDM6B expression is lost. Breast cancer is another example where progesterone can increase levels of EZH2 to promote mammary epithelial cell proliferation, and excessive EZH2 expression can occur in cancer. This exemplifies how epigenetics can integrate environmental signals and have a profound influence on the fine balance between stem cell maintenance and overt carcinogenesis. In general, a recurrent theme in cancer is loss of key chromatin regula tion that promotes cell differentiation, combined with gain of activities that promote stemness. Chromatin-modifying enzymes also play a major role in influenc ing cell-type specificity. High levels of EZH2 that modify H3K27me3 promote adipogenesis while simultaneously inhibiting osteogenesis. In contrast, the H3K27me3 demethylases, KDM6A (UTX) and KDM6B, derepress those same genes, driving stem cells toward osteogenesis. Through interactions with tissue-specific master regulators, epigenetic modifiers also shape cell-type specificity. In the epidermis, p63, the p53 family member that is a master regulator of the epidermal compart ment, interacts with several chromatin regulators including HDAC1 and HDAC2, SATB1, MLL4 (KMT2D), and BRG1 to orchestrate epidermal differentiation. Similarly, the gene-activating H3K4 histone methyltransferases, MLL3 (KMT2C) and MLL4, are required for

adipogenesis by forming a complex with the transcriptional activator ASC2 and the transcription factor PPARγ to induce adipogenic genes. Overall, loss of epigenetic regulation can reduce cell differentiation and increase stem cell specification to drive diseases encompassing devel opment, cancer, and, broadly, diseases associated with aging.

■ ■EPIGENETICS OF METABOLISM One of the fascinating aspects of epigenetics is that it represents a mechanism for direct connection between the environment and gene expression. Numerous studies in the field of metabolism have identi fied a complex interplay between diet, metabolism, and the epigenome (Fig. 497-1). Seminal findings in Drosophila and mice have shown that changes in diet, particularly the paternal diet, and other environmental factors, can influence the metabolism of offspring, ultimately promot ing obesity in later generations. Epidemiologic studies in humans have supported these results, as the nutritional status of grandparents has been correlated with phenotypic effects in grandchildren. In fact, diet can directly affect the levels and activity of chromatin modifiers. For instance, high-fat diets reduce histone acetylation through their ability to inhibit the enzymes ACLY and ACSS2, which produce acetylCoA. Levels of acetyl-CoA, in comparison to all measured metabolites, are indeed the best predictor of histone acetylation levels. Consistent with this, increased acetyl-CoA correlates with rising levels of total histone acetylation, including at the promoters of growth-associated genes. This increase in nuclear acetylation is associated with cell cycle progression and proliferation, and it can have clinically relevant down stream effects. For example, high levels of acetyl-CoA can delay stem cell differentiation and suppress autophagy. The oncogenes MYC and AKT can both hijack metabolic networks to enhance nutrient uptake by cancer cells, thus promoting acetyl-CoA production and resulting in both the initiation and progression of tumorigenesis. Additional evidence suggests that dietary intake of alcohol can directly contribute to acetate levels and therefore histone acetylation in the brain, with effects on the transcription of genes involved in learning and memory. CHAPTER 497 The Role of Epigenetics in Disease and Treatment Contrary to convention that metabolic enzymes are strictly mito chondrial or cytosolic, certain metabolic enzymes can be present in the

nucleus and can thereby directly regulate histone acetylation enzymes. This is the case for several enzymes that generate acetyl-CoA, includ ing ACLY, PDH, and ACSS2, which generate acetyl-CoA from citrate, pyruvate, and acetate, respectively. Further, ACSS2 can be chromatinbound to regulate gene expression, leading to physiologic responses such as autophagy in the liver and mammalian hippocampal func tion in learning. This direct metabolic-epigenetic enzyme cross-talk illuminates a crucial local role of the acetyl-CoA metabolite to effect rapid gene transcription and represents a fertile intersection for future therapeutics. Methylation is also altered by metabolism. S-Adenosylmethionine (SAM) is the key metabolic cofactor for histone and DNA methyla tion. Dietary factors are estimated to explain 30% of the variation in human serum methionine concentration and hence can alter SAM levels and histone methylation. For example, dietary methionine avail ability and intracellular production of SAM affect the levels of histone H3K4me3 associated with transcriptional activation. Furthermore, these fluctuations can have critical physiologic consequences: DNA methylation levels in rectal mucosa and colonic polyps are increased by higher levels of dietary folate, and a diet low in methyl donors reduces the formation of gastrointestinal cancers in mice predisposed to these tumors. Methionine metabolism and the availability of SAM regulate stem cell differentiation and contribute to carcinogenesis. For instance, cancers with mutations in metabolic regulatory genes such as IDH1/2, FH, and SDH lead to the accumulation of by-products (2-hydroxy glutarate, fumarate, and succinate, respectively), which all inhibit α-ketoglutarate (α-KG)–dependent histone demethylases and thus pro mote hypermethylation and lead to impaired cellular differentiation. Notably, some of the α-KG–dependent demethylases, which are highly mutated in numerous cancers (i.e., KDM5A, KDM6A), also serve as cellular oxygen sensors, thus linking environmental oxygen levels to epigenetic control of methylation levels. In contrast to hypermethyl ated states, loss of the MTAP gene, which is part of the 9p21 locus

containing p16 and one of the most frequent events in human cancer, disrupts normal methionine metabolism. This both lowers methyla tion levels, and, interestingly, also sensitizes cancer cells to inhibitors of the PRMT5 methyltransferase, therefore opening a therapeutic opportunity. These observations illustrate how connections between epigenetics and metabolism can generate unanticipated advances in medicine. Furthermore, these data highlight the tight interconnections between environmental inputs, metabolism, and epigenetics.

■ ■CANCER EPIGENETICS Cancer is now understood to be a mixed genetic and epigenetic disease, as epigenetic dysregulation is pervasive in human cancers (Fig. 497-1). Beyond simple activation of oncogenes or reduced expression of tumor suppressors, epigenetic mechanisms can contribute to chemotherapy resistance and to failure of antitumor immunity. Accordingly, the development of drugs targeting epigenetic pathways is one of the most active areas of clinical and pharmaceutical development, with several compounds already approved for human use and shown to be effec tive in a variety of cancers. Epigenetic perturbations in cancer largely affect chromatin-regulating enzymes, which represent robust targets for development of novel small-molecule inhibitors, especially as com pared with canonical oncogenic transcription factors (e.g., MYC) and tumor suppressors (e.g., p53). Epigenetics can contribute to carcinogenesis in a variety of ways. First, on a global scale, chromatin organization is the single most influ ential factor in determining local mutation rate across the genome. Analysis of abundant tumor sequencing data has demonstrated that heterochromatic regions of the genome contain a higher frequency of mutations compared with more open euchromatic regions. This differ ence is due to the improved accessibility of the

DNA repair machinery to less compact, more open regions of chromatin. PART 20 Emerging Topics in Clinical Medicine The first discovery of an epigenetic mutation was found in 1998 when the chromatin remodeler SMARCB1 was shown to drive the for mation of malignant rhabdoid tumors. Extensive sequencing of human tumors from the majority of cancer types has been performed by The Cancer Genome Atlas (TCGA) consortium, and remarkably, 25–30% of identified cancer driver mutations occur in chromatin regulatory proteins. Similar to SMARCB1, numerous other chromatin modifi ers (e.g., methyltransferases MLL3 and MLL4, and acetyltransferases EP300 and CBP) and nucleosome remodeling enzymes and associated complex components (e.g., SMARCA4, SMARCA2, ARID1A) are heav ily mutated and inactivated in many cancers. The majority of these mutations are loss-of-function mutations, and indeed, enzymes like MLL4 and demethylase KDM6A possess tumor-suppressive activ ity across a variety of tissues and cellular contexts. In contrast, the H3K27me3 histone methyltransferase EZH2 is an oncogene, and accordingly, it is overexpressed in many advanced-stage or metastatic solid tumors such as breast cancer, prostate cancer, and melanoma. Mechanistically, EZH2 represses the p16 tumor suppressor and other cell cycle genes required for cell cycle exit via H3K27me3 deposition. Consistent with a broad growth regulatory role, EZH2 inhibitors are therapeutically successful for a number of cancers in preclinical models and are being actively studied for B-cell lymphoma, melanoma, and other solid tumors. In addition, provocative evidence has emerged for a direct tumori genic role of histones based on the discovery of causative mutations, such as histone H3 mutations identified in pediatric high-grade glio mas. Specifically, the majority of these mutations are in the H3 variant H3.3, where lysine 27 is replaced by methionine (K27M). Similarly,

> 90% of chondroblastomas replace lysine 36 with methionine (K36M) in histone H3.3. These effects appear to be dominant negative because (1) in H3.3, these are heterozygous mutations, and (2) the mutations also occur in the canonical H3, which exists in ~30 orthologous genes in the human genome. Thus, a minority of H3/H3.3 mutant protein leads to global defects in the associated histone modifications (K27 or K36 methylation), possibly via irreversible inhibition of the cognate enzymes by the mutant histones. These “oncohistone” mutations pro mote resistance to apoptosis and failure of normal differentiation in a number of pediatric and adult cancers.

Beyond mutations, genetic translocations involving chromatin modifiers also implicate chromatin pathways as direct drivers in cancer. MLL1 (KMT2A), the H3K4 histone methyltransferase, is a frequent translocation partner occurring in adult and pediatric acute myeloid leukemia (AML) and in ~80% of infant acute lymphoid leuke mia (ALL) cases. MLL1 can fuse with >70 translocation partners, and these mutant proteins prevent normal hematopoietic differentiation. Consistent with a causative role of MLL1 in these gene fusions, drugs inhibiting the catalytic activity of MLL1 are effective in preclinical models of AML and are currently being evaluated in human clinical trials. Given the abundance of epigenetic abnormalities in cancer com bined with the inherent reversibility of epigenetic changes, extensive efforts are underway to develop epigenetic drugs. The first epigenetic therapeutic involved the use of DNA methylation inhibitors (DNMTi) to reactivate tumor-suppressor genes. Interestingly, the mechanism of traditional chemotherapeutics, such as azacitidine and decitabine, is to inhibit DNMT1, thereby promoting global hypomethylation;

these are currently in clinical use for myelodysplastic syndrome (MDS) and AML. In a second broad mechanism, loss of acetylation occurs in many cancers, and thus, HDAC inhibitors (HDACi) are under inten sive development. HDACi are effective and approved for treatment in cutaneous T-cell lymphoma and multiple myeloma. Bromodomain (BRD)-containing proteins bind to lysine acetylated target proteins, including histones, and rationally designed BET inhibitors (BETi) block their binding. BETi reduce the amplified expression of oncogenes such as MYC in hematologic cancers. Current studies are now focused on optimizing combinatorial epigenetic therapies with conventional chemotherapies and immunotherapies, particularly given the ability of epigenetic therapeutics to promote re-expression of tumor antigens and interferon (IFN)-mediated antitumor immunity. Indeed, the development of a new generation of more specific epigenome-targeted inhibitors, combined with our increased knowledge of the underlying epigenetic mechanisms contributing to tumorigenesis, has enabled a precision medicine–based approach to harnessing the potential of these drugs. This may be particularly valuable in the context of improv ing patient responses to a variety of therapies beyond chemotherapies and immunotherapies, such as radiation and hormone therapies. There are several hundred chromatin enzymes and binding proteins in the human genome, and the current focus is to identify more specific inhibitors. Indeed, targeted inhibitors of numerous mutated chroma tin regulators have been developed, with >30 compounds currently in various stages of development and preclinical trials. Some notable examples showing early clinical success include EZH2 inhibitors for lymphomas, sarcomas, and melanoma; IDH inhibitors for AML and gliomas carrying mutant IDH1 or IDH2 genes; LSD1 inhibitors for AML and small-cell lung cancer; and DOT1L and MLL1 inhibitors for leukemias with activated MLL1. Given the broad potential effects of epigenetic regulators, it is perhaps not surprising that there have been some dose-limiting toxicities, particularly among those that are less target-specific. Collectively, the emerging picture is that the most effec tive and robust use of epigenetic drugs in cancer will be fine-tuning and potentiating the effects of other therapies that are either incom pletely effective or marked by widespread resistance. ■ ■EPIGENETICS OF AGING Like many diseases of aging, human aging itself results from the complex interplay between genes and the environment. Evidence that the epigenome may be the key link between these processes derives from observations that numerous environmental stimuli and stressors—ranging from diet and exercise to hormones and circadian rhythms—contribute to both aging and epigenetic alterations (Fig. 497-1). Thus, a lifetime of exposures progressively disrupts the chro matin landscape. These age-dependent changes in chromatin organi zation increase the susceptibility of the genome to mutations and also reduce transcriptional fidelity. Further, provocative findings in model systems demonstrate that stress-induced epigenetic changes can be transmitted over several generations and can even affect the life span of later generations. Among these global epigenetic alterations, there

is dysregulation of histone modifications and a general loss of histone proteins with aging across taxa. Amazingly, experimental increases in histone levels, particularly histones H3 and H4, but not H2A or H2B, can reverse these age-related changes in mammalian cells and in the yeast Saccharomyces cerevisiae model. Thus, the sum of current evidence suggests a model of aging via a general increase in activating epigenetic modifications along with a loss of repressive modifications. Together these changes create a state of transcriptional instability and “noise” that inhibits accurate transcrip tion. Cells from patients with Hutchinson-Gilford progeria syndrome (HGPS), the most severe form of human premature aging, display reduced levels of both H3K9me3 and H3K27me3 repressive chromatin. In another premature aging disease, Werner syndrome, DNA

damage induces global loss of H3K9me3 and H3K27me3 due to the inherent absence of the Werner syndrome ATP-dependent DNA helicase, which is critical for DNA repair. Such heterochromatin loss is not limited to premature aging conditions, as aged cells derived from healthy older humans display age-dependent loss of H3K9me3 leading to aberrant expression of normally repressed transposable elements. Activation of these mobile elements correlates with neurodegenerative disorders and may also promote other aging-related phenotypes such as cancer. Human fibroblasts undergoing cellular senescence (exit from cell cycle due to replicative or other stress) undergo destabilization of compact heterochromatin adjacent to the nuclear periphery, in so-called laminassociated domains (LADs). At LADs, in addition to a reduction of repressive histone modifications as discussed above, there are broad new regions of the euchromatic histone modification H3K4me3. This general loss of heterochromatin can promote the activation of cytosolic DNA and RNA sensing pathways that promote innate immune signal ing and "inflammaging." In addition to age-associated alterations of histone modifications, direct manipulation of chromatin-modifying enzymes that control these marks affects the balance between heterochromatic and euchro matic regions, and it alters the lifespan of model organisms. Inhibiting the H3K27me3 histone demethylase KDM6A results in increased repressive H3K27me3 and extended lifespan in Caenorhabditis elegans. Consistent with this, genetic reduction of enzymes (ash-2, set-2, wdr-5) that add the activating H3K4me3 histone modification also extends lifespan in C. elegans. The consequences of these genetic manipulations nicely correspond to the observed changes in histone modifications as described above. Beyond histone-modifying enzymes, dysregulation of the levels or function of chromatin remodelers can also affect lifespan in model organisms. This dysregulation occurs in humans as well, as in the nucleosome remodeling deacetylase complex (NuRD), which is reduced in HGPS fibroblasts and in aged healthy donors. In addition to age-related changes in histone methylation, histone acetylation also contributes to aging phenotypes. Dysregulation of his tone acetyltransferases (HATs) and HDACs is associated with reduced longevity across model organisms. Further, sirtuin deacetylases (class III NAD+-dependent HDACs) promote health span and lifespan across species as key mediators of pro-longevity effects of caloric restriction. Indeed, loss of Sirt6 results in premature aging in mice, while caloric restriction–induced increases of Sirt1 and Sirt6 expression can delay aging. As discussed previously, metabolism and acetylation are intri cately linked, and the sirtuins, via NAD+ levels, and other HDACs may play key roles connecting the environment, gene expression, and physiologic output. For instance, exercise in humans reduces activity of HDACs 4 and 5, leading to increased H3K36ac in skeletal muscle, which likely promotes beneficial gene expression. Epigenetic alterations with aging are not limited to histone modifi cations and extend to DNA methylation. Consistent with the histone patterns, DNA methylation data support the model described above—that is, general decompaction of the epigenome with aging. Specifically, levels of 5-mC are reduced in senescent human cells, and global DNA hypomethylation occurs across the human genome with aging. Con current with this overall hypomethylated state, there are local regions of hypermethylation focused near CpGs at gene promoters, particu larly at genes that maintain cellular differentiation and cell identity. This epigenetic disruption during aging thus leads to profound changes

in transcription. For example, in hematopoietic stem cells, DNA hyper methylation blocks proper binding of transcription factors, resulting in dysregulation of normal gene expression with aging. Importantly, these patterns are not merely stochastic alterations in response to environmental stressors throughout aging. Indeed, the methylation status of a defined number of CpG sites is a

highly accurate predictor of chronologic age in human tissues. This work reveals that DNA meth ylation status with aging outperforms previous standard biomarkers of aging, such as p16 expression levels and telomere length, and will be highly valuable in the near future to gauge effects of treatment aiming to ameliorate diseases of aging.

■ ■EPIGENETICS OF THE BRAIN AND BEHAVIOR Brain disorders are among the greatest clinical challenges to under stand and to treat. Most neurologic and psychiatric disorders result from complex dysregulation of numerous genes and pathways. In this interplay between underlying genetic predisposition and external environmental factors, aberrant epigenetic regulation is increasingly recognized as a potentially key modulator (Fig. 497-1). More directly, however, several progressive neurodevelopmental disorders are caused by germline mutations in chromatin regulators. Mutations in methyl CpG binding protein 2 (MECP2), a protein important for binding to methylated DNA and contributing to gene repression, are the major cause of Rett syndrome. MeCP2 loss leads to overactive gene transcription in neurons and impaired presynaptic excitatory functions. Similarly, Kabuki syndrome, another progres sive neurodevelopmental disorder, is caused by germline mutations in either the H3K4me1 histone methyltransferase, MLL4 (KMT2D), or the H3K27me3 demethylase, UTX (KDM6A). This disorder may derive from dysregulation of transcriptional enhancers, a major class of gene regulatory elements, as both MLL4 and UTX play a key role in activa tion of enhancers. Finally, the acetyltransferase CBP (CREBBP) also is important for gene enhancer function and, when mutated, can lead to Rubinstein-Taybi syndrome, a cause of intellectual disability. CHAPTER 497 The Role of Epigenetics in Disease and Treatment Beyond germline mutations, altered methylation dynamics can drive disorders of neural development and of neurodegeneration. Fragile X syndrome, characterized by learning disabilities and cogni tive impairment, is caused by mutations in the FMR1 or FMR2 gene or by hypermethylation of the transcriptional promoters regulating FMR1 or FMR2. Similarly, Prader-Willi syndrome and Angelman syndrome, neurodevelopmental conditions caused by abnormal imprinting of the paternal or maternal chromosomal region (15q11-13), respectively, are frequently caused by aberrant DNA methylation. Further, DNA hypomethylation is implicated in some neurodegenerative condi tions. For instance, in Parkinson’s disease, several genes involved in pathogenesis are hypomethylated due to DNMT1 depletion, includ ing the α-synuclein gene (SCNA). In Alzheimer’s disease (AD), DNA hypomethylation occurs at promoters of key pathogenic genes such as amyloid precursor protein (APP). Indeed, APP promoter methylation is responsive to environmental factors, including aging, a major risk factor for AD. Likewise, presenilin-1 (PSEN1) is implicated in AD and displays altered DNA methylation in response to variations in metabolic stimuli. Recent evidence from human AD brains demon strated significant enrichment of H3K9 and H3K27 acetylation and provided evidence that this dysregulation of the epigenome promotes gene transcription pathways involved in AD pathogenesis. Studies of Huntington’s disease (HD) have demonstrated DNA hypomethylation and decreased histone acetylation, in part due to altered function of the acetyl transferase CBP, leading to transcriptional dysregulation. Together, these observations underscore altered epigenetic regulation as a crucial feature of neurodegeneration. Additional gene regulatory proteins in the nervous system interact with and are regulated by chromatin modifiers. REST (repressor ele ment 1–silencing transcription factor) is important in neuronal homeo stasis through its ability to recruit chromatin regulatory enzymes, such as histone deacetylases and histone methyltransferases, and via its control over gene expression. REST levels increase with aging and serve a protective function in neurons against age-associated stressors and loss of cognitive function associated with AD. Similar to REST,

brain-derived neurotrophic factor (BDNF), another important media tor of neural development and homeostasis, is implicated in a variety of neurologic and psychiatric disorders including HD, depression, schizophrenia, bipolar disorder, and autism. Knockdown of BDNF in the dentate gyrus leads to depression-like behavior in mouse models, and BDNF mediates effects of antidepressant therapies. Chromatin pathways, including DNA methylation/MeCP2 and H3K27me3, play a key role in BDNF regulation as observed in brains from patients with schizophrenia.

Finally, addiction medicine is another frontier where epigenetics holds great promise to reveal connections between environmental exposure and phenotypes. Although still in its early stages in terms of mechanistic understanding, emerging evidence demonstrates disrup tion of epigenetic homeostasis as a consequence of addictive substances ranging from alcohol to cocaine. For example, the acetylation of regula tory elements in the FOSB gene by the histone acetyltransferase CBP is associated with behavioral effects of cocaine. Opioid exposure appears to promote a generally more open and permissive state of chromatin marked by increases in histone acetylation and reductions in histone methylation, which may allow for a more hyperresponsive state and reinforce reward-seeking behaviors. Ethanol also induces histone acetylation and a decompacted chromatin structure with direct effects on learning and memory function. ■ ■EPIGENETIC INFLUENCES ON INFECTION, IMMUNITY, AND INFLAMMATION Alterations in gene expression patterns are important determinants of immune-mediated disease, and in turn, epigenetics regulates infection, immunity, and inflammation (Fig. 497-1). Treatment with immunestimulating agents such as lipopolysaccharide (LPS) and tumor necrosis factor α activates expression of numerous inflammatory genes within hours, with precise gene pathways and activation kinetics determined by the cellular epigenetic state. HATs and HDACs are critical components of this response, coordinating with proinflammatory transcription fac tors, such as AP-1 and NF-κB, to either activate (in the case of HATs) or repress (in the case of HDACs) inflammatory genes. For example, cor ticosteroids recruit HDAC2 to promoters of NF-κB–stimulated inflam matory genes to prevent activation during asthma treatment. PART 20 Emerging Topics in Clinical Medicine Type 1 IFN responses are exceptional examples of regulatory com plexity governed by epigenetic control. In an unstimulated state, the H3K9 methyltransferases G9a (EHMT2) and EHMT1 suppress expres sion of IFN and IFN-induced genes. Upon induction of IFN-stimulated genes, STAT transcription factors recruit chromatin remodeling com plexes, such as BAF (SMARCA4), and recruit HATs including p300, CBP, and GCN5 (KAT2A). In turn, chromatin remodeling and acetyla tion recruit chromatin binding proteins including the bromodomain protein, BRD4, which promotes transcriptional elongation and full activation. Beyond the DNA level, METTL3-mediated m6A methyla tion on mRNAs also is a critical regulator of IFN signaling in a variety of distinct cellular contexts. Major regulators of adaptive immunity pathways are similarly epige netically regulated. CD4+ and CD8+ T cells undergo extensive changes in histone modification profiles during differentiation to distinct subsets of effector T cells. For example, genes associated with effector T-cell functions in CD8+ memory T cells (e.g., PRDM1, KLRG1, IFNG) display enrichment of H3K4me3 and low levels of H3K27me3 com pared with those genes in naïve T cells. DNA methylation also plays an important regulatory role and may contribute to disease. For example, CD4+ T cells from individuals with rheumatoid arthritis (RA), sys temic scleroderma, and latent autoimmune diabetes in adults display hypermethylation of the FOXP3 gene, which activates regulatory T cells that dampen immune responses. In addition, hypermethylation of the CTLA4 locus occurs in regulatory T cells from RA patients, impair ing their immunosuppressive abilities. During infection, epigenetic processes can play critical roles in both the immune response and defense against pathogens, as

well strategies exploited by microorganisms to co-opt the host cellular machinery to advantage of the pathogen. Respiratory syncytial virus (RSV) infection promotes the expression of the histone demethylase KDM5B, which

removes H3K4 methyl groups from antiviral genes such as type 1 IFNs, driving a switch from T helper 1– to T helper 2–type immune responses, thereby contributing to chronic infection. Similarly, influ enza upregulates the repressive H3K9me3 methyltransferase SETDB2 to block expression of CXCL1 and a variety of NF-κB target genes involved in attracting neutrophils and host defense, both serving to lengthen the infection and contributing to bacterial superinfection. Regarding the host response to infection, studies have revealed that differences in host tissue-, age-, and sex-biased epigenetic profiles might shape susceptibility and responses to infection. For example, differential DNA methylation at the ACE2 gene may impact expression levels of this key cellular receptor and ultimately the ability of SARSCoV-2 to infect hosts, while alterations in antiviral IFN signaling may lead to more severe COVID-19 infection and disease. These findings are all supported by new discoveries demonstrating that epigenetics is a key component for the inflammatory memory that has been observed now across a wide variety of contexts. Numerous perturbations ranging from infections and vaccination to skin wounding and Western diets have now been shown to elicit an epigenetic memory that is maintained and propagated. This epigenetic memory extends beyond just the immune system to the involved tissues. These findings have suggested a potential for epigenome-modifying drugs for the treatment of inflammatory and immune-related condi tions. For example, the DNA methylation inhibitors azacitidine and decitabine have immunosuppressive effects possibly mediated by enhanced expression of FOXP3, which generally suppresses immune responses. HDACi upregulate and downregulate immune genes, and they inhibit cytokine production in macrophages from patients with RA. Further, the HDACi vorinostat and panobinostat inhibit primary B-cell responses and antibody production in vitro and in vivo. Given these broad effects, it is not surprising that the HDACi trichostatin A (TSA) has efficacy in various model systems for treatment of RA, systemic lupus erythematosus (SLE), asthma, acute kidney injury, sepsis-induced lung and cardiac damage, and acute pancreatitis. Simi larly, BETi also display broad effects in blocking antigen presentation and T- and B-cell activation and thus beneficial protective effects in a variety of inflammatory settings including autoimmunity, sepsis, atherosclerosis, psoriasis, periodontitis, and arthritis. Beyond these “broad-spectrum” epigenetic inhibitors, GSK-J4, which is a specific inhibitor of the H3K27me3 demethylases KDM6A and KDM6B, has anti-inflammatory activity, presumably by preventing loss of H3K27me3 repression over inflammatory genes. Similarly, inhibition of the H3K4me3 histone methyltransferase MLL1 blocks the induction of proinflammatory cytokine gene expression in a variety of contexts. CONCLUSION Due to the enormity and complexity of the chromatin and epigenetics fields and their reach into all areas of biology and medicine, it is not possible to cover such a broad scope in a single chapter. Thus, here we provide a concise snapshot highlighting key areas of development in medicine. We hope to have conveyed the tremendous excitement and promise that pervades the discipline. Indeed, given the expo nential growth in uncovering the interface between the epigenome and epigenetic therapies with the environment and disease, there is little doubt that the coming years will bring important additions to this field. ■ ■FURTHER READING Bates SE: Epigenetic therapies for cancer. N Engl J Med 383:650, 2020. Carter B, Zhao K: The epigenetic basis of cellular heterogeneity. Nat Rev Genet 22:235, 2021. Dai Z et al: The evolving metabolic landscape of chromatin biology and epigenetics. Nat Rev Genet 21:737, 2020. Dinardo AR et al: Postinfectious epigenetic immune modifications— a double-edged sword. N Engl J

Med 384:261, 2021. Hwang JY et al: The emerging field of epigenetics in neurodegenera tion and neuroprotection. Nat Rev Neurosci 18:347, 2017. Janssen SM, Lorincz MC: Interplay between chromatin marks in development and disease. Nat Rev Genet 23:137, 2021.

06 - 498 The Role of Circadian Biology in Health and Disease

498 The Role of Circadian Biology in Health and Disease

Millán-Zambrano G et al: Histone post-translational modifica tions—cause and consequence of genome function. Nat Rev Genet 23:563, 2022. Sendinc E, Shi Y: RNA m6A methylation across the transcriptome. Mol Cell 83:428, 2023. Vandereyken K et al: Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet 24:494, 2023. Zhang W et al: The ageing epigenome and its rejuvenation. Nat Rev Mol Cell Biol 21:137, 2020.

The Role of Circadian

Biology in Health

and Disease Jonathan Cedernaes, Kathryn Moynihan Ramsey, Joseph Bass Circadian rhythms are anticipatory, circa 24-h, autonomous cycles of physiology and behavior. These evolutionarily conserved rhythms have evolved at both the cell and tissue level to synchronize organismal function in anticipation of the 24-h rotation of the Earth. A common feature of modern “24/7” life is the routine disruption of these endog enous circadian cycles due to the rise in shift work, jet travel across time zones, exposure to blue light–emitting devices at night, and dis rupted sleep-wake behavior. In-depth characterization of the molecular basis of circadian disorders has generated novel avenues for research on how sleep-wake disruption has been associated with aging, metabolic disease, inflammation, and cancer. This chapter provides an overview of (1) the basic biology of the circadian system; (2) primary circadian rhythm and interrelated sleep disorders; and (3) the role of the circa dian system in both normal human physiology and disease states. We also include an overview of how the emerging field of chronobiology may impact drug action. A glossary of terms used in circadian biology is summarized in Table 498-1. ■ ■BASIC EVOLUTION AND STRUCTURE

OF THE CIRCADIAN SYSTEM Long before the emergence of multicellular life, the Earth’s constant rotation around its axis gave rise to a daily cycle of light and darkness. At the emergence of the

first prototypal gene involved in biological clock regulation—3.4 billion years ago in photosynthetic cyanobacteria— the period of Earth's rotation along its own axis was only 8 h. The co-occurrence in molecular evolution of the biological clock and photosynthesis hints at an interrelated and selective advantage of the clock in the regulation of energetic processes. Indeed, biological clocks coordinate oxygenic reactions with periods of sunlight each day, and perturbation of clock cycles reduces fitness, reproduction, and survival. Additionally, clocks protect photosynthetic organisms from the DNAdamaging effects of sunlight by timing the production of DNA repair processes, such as photolyase-mediated repair, to the nighttime. Across billions of years of evolution, as day length has gradually extended to today's circa 24 h, highly conserved circadian clocks (from circa diem, meaning "about a day") have been found in all photosensitive organisms, governing a wide range of biochemical, physiologic, and behavioral processes. A defining property of the circadian clock system is that it enables organisms to anticipate, rather than simply react to, daily changes in the external environment that are tied to the day-night cycle. In mammals, circadian systems are organized hierarchically with a light-responsive "master" circadian pacemaker located within the suprachiasmatic nucleus (SCN) of the anterior hypothalamus, which in

turn presides over a network of both extra-SCN and peripheral clocks (see "Anatomic Organization of the Circadian Clock Network" below). Daily light exposure signals to the SCN and entrains the circadian system to the 24-h day (see "Entrainment and Measurement of the Circadian System," below). In turn, the SCN maintains synchrony of a diverse network of both central and peripheral clocks via a variety of signals that have as of yet to be fully identified. These signals involve direct physiologic rhythms (core body temperature), the autonomic nervous system, and neuroendocrine signals, such as cortisol, which is part of the hypothalamic-pituitary-adrenal (HPA) axis.

■ ■MOLECULAR ORGANIZATION OF THE MAMMALIAN CIRCADIAN CLOCK At the molecular level, mammalian circadian rhythms are generated by a transcription-translation autoregulatory feedback loop. The forward limb of the clock is composed of the basic helix-loop-helix transcrip tion factors (TFs) CLOCK (or its paralogue, NPAS2) and BMAL1. These drive expression of their own repressors (PERs and CRYs) in the negative limb in a cycle that repeats itself every 24 h (Fig. 498-1). A second short feedback loop involves CLOCK/BMAL1-mediated tran scription of the retinoic acid–related orphan nuclear receptor families ROR and REV-ERB, which activate and repress Bmal1 transcription, respectively. Rhythmic posttranslational regulation of the stability and degradation of core clock TFs occurs via events such as phos phorylation by casein kinase 1 epsilon (CK1ε) and casein kinase 1 delta (CK1δ) and ubiquitination by FBXL3 and FBXL21. In addition to the circa 24-h oscillation of core clock genes, a wide array of downstream clock-controlled genes (CCGs) exhibit broad rhythmic amplitude in expression, ultimately giving rise to rhythmic physiologic processes. CHAPTER 498 The importance of localized clock gene expression has been dem onstrated by genetic animal studies, such as with targeted ablation of Bmal1, the only clock gene that lacks a known functional paralog. Deletion of Bmal1 either in the whole brain or in regions that span the brain region that coordinates circadian rhythms—the SCN—causes behavioral arrhythmicity, even when genetic ablation occurs in adult life. Conversely, restoring Bmal1 expression specifically in brain in global adult Bmal1 mutant mice rescues behavioral locomotor rhythms. Of note, whereas the protein CLOCK normally heterodimer izes with BMAL1, the paralogous protein NPAS2 can functionally substitute for CLOCK within the pacemaker neurons. Thus, while mice lacking either Clock or Npas2 genes maintain rhythmicity, mutants lacking both CLOCK and NPAS2 lack circadian rhythms in locomotor activity. Further, mutations in many of the clock genes are associated with

impaired circadian rhythms and physiology in both experimental animal models and humans (see “Primary Pathologies of the Circadian System” below). The Role of Circadian Biology in Health and Disease

A major transformation in our understanding of circadian biology came with the discovery that the molecular clock network is present not only in the SCN but also within most peripheral tissues, as well as in extra-SCN neurons in the brain. In primates, ~82% of all proteincoding transcripts exhibit daily 24-h rhythms in some tissue or other. In rodents studied under constant conditions, ~3–16% of the transcrip tome in each tissue exhibits 24-h rhythms in mRNA expression levels, even though the repertoire of such genes varies substantially between tissues, in accordance with tissue-specific functions. The core clock feedback loop and the induction of transcriptional CCG rhythms also involves epigenetic mechanisms such as conformational chromatin dynamics, histone acetylation, and DNA methylation. Conversely, posttranscriptional events such as RNA polyadenylation, nucleocy toplasmic shuttling, alternative splicing, and mRNA translation also exhibit circadian variation, further increasing the repertoire of rhyth mic regulation at a cellular level. ■ ■ANATOMIC ORGANIZATION OF THE

CIRCADIAN CLOCK NETWORK The molecular circadian feedback loop is synchronized with sunrise each day by photosensitive melanopsin-expressing neurons within the retina. These neurons provide input to the SCN via the retino hypothalamic tract (RHT), allowing mammals to maintain coherent

TABLE 498-1 Glossary of Terms Used in Discussion of the Circadian System TERM DESCRIPTION ASPD Advanced sleep phase disorder (see text for description). CBT Core body temperature. Often used as an indicator of the circadian rhythm but can be masked by sleep and exercise. CCGs Clock-controlled genes; output of the molecular clock. Chronotype Internal circadian rhythm of an individual determined by phase of entrainment, determining sleep propensity and timing of maximum alertness over a 24-h period. Circadian period Time required for one complete cycle or oscillation. Calculated by the time distance between two consecutive peaks or troughs of a circadian variable. Circadian phase Timing of the circadian rhythm. Defined by comparing, e.g., the peak (acrophase) or trough (bathyphase) to a fixed event, e.g., to a point in time. Synonymous with phase angle. Circadian rhythm A biological process that exhibits an endogenous, entrainable oscillation of ~24 h. Circadian rhythm sleep disorders Disorders of multiple etiology that have in common that they result in maladjustment of the biological clock with respect to the environment. Constant routine An experimental paradigm designed to study endogenous circadian rhythms in humans, by keeping behavioral and environmental factors constant. These paradigms thereby typically entail a combination of constant dim lighting, evenly distributed isocaloric energy intake, semirecumbent posture, and forced extended wakefulness. Desynchrony Loss of synchrony occurring either between a rhythm and its zeitgeber (external, “time giver” signal) or between two or more rhythms within an organism (internal). Diurnal rhythm An oscillation synchronized with the day/night cycle that repeats itself with a 24-h period. The rhythm does not have to persist when time cues (e.g., light) are absent. DLMO Dim-light melatonin onset; a marker of melatonin rhythm. DSPD Delayed sleep phase disorder (see text for description). Entrainment Synchronization of a circadian rhythm or other self-sustaining oscillation by a factor—zeitgeber—that enforces the oscillator. Constant entrainment between the zeitgeber and the oscillator results in a stable phase relationship between these entities. Infradian rhythm A recurrent cycle or period with a period length significantly greater than 24 h. Melatonin Hormone produced primarily by the pineal gland

(chemical name N-acetyl-5-methoxytryptamine); derived from L-tryptophan. Various forms of melatonin can be prescribed for circadian rhythm sleep disorders or sleep disorders. PART 20 Emerging Topics in Clinical Medicine Non-24-h rhythm disorder A syndrome in which there typically are chronic 1- to 2-h daily delays in sleep onset and wake times in an individual living in society, e.g., due to complete blindness. Peripheral clocks Clocks presiding outside of the suprachiasmatic nucleus, the circadian system's master pacemaker. PRC Phase response curve; visual representation of how a particular manipulation (e.g., light) produces phase shifts as a function of the phase (i.e., circadian time) at which the manipulation occurs. Defining the PRC to light has enabled researchers to understand and predict how entrainment to light cycles is accomplished. SCN The suprachiasmatic nucleus or nuclei, also known as the master pacemaker in mammalian species. A bilateral set of nuclei positioned in the anterior ventral hypothalamus. Essential for entraining extra-SCN central and peripheral oscillators to the prevailing light-dark cycle via photic input from the retina. Shift work Work scheduled so that it occurs outside of the traditional work schedule of 9:00 a.m. to 5:00 p.m., or 7:00 a.m. to 6:00 p.m., depending on definition. Various forms of shift work exist, such as early morning, evening, or night shifts, as well as rotating shifts. Ultradian rhythm A recurrent cycle or period with a period significantly shorter than 24 h—e.g., a 2-h rhythm would exhibit 12 cycles within a circadian (24-h) rhythm. RRE Bmal1 CK1ε/δ PERs P CRYs P CLOCK BMAL1 E-box RORα/γ FIGURE 498-1 Central clock molecular mechanism. The core molecular clock machinery in mammals is encoded by interlocking transcription-translation feedback loops that oscillate with ~24-h periodicity. The transcription factors CLOCK and BMAL1 heterodimerize to drive transcription of downstream clock-controlled target genes containing E-box enhancer elements. Among these, the PER and CRY proteins multimerize and inhibit CLOCK/BMAL1, while RORs and REV-ERBs activate and inhibit, respectively, Bmal1 transcription, resulting in rhythmic oscillations of clock-controlled and downstream target genes.

Stabilization FBXL21 CRYs CRYs P FBXL3 Degradation Clock-controlled genes Rev-erbα/β

Environmental inputs and internal circadian organization Brain Clocks SCN Environmental light/dark cycle SCN Extra-SCN LHA PVN ARC PIT Non-autonomous circadian control Peripheral Clocks Vasculature Liver Adrenals Pancreas Muscle Environmental nutrient cycle fasting/feeding Fibroblasts Intestine Hematopoetic Adipose Autonomous circadian control FIGURE 498-2 Central and peripheral clocks coordinate environmental cues with behavior and physiologic outputs. Light entrains the master pacemaker neurons in the suprachiasmatic nucleus (SCN), which subsequently synchronizes extra-SCN and peripheral clocks. Brain clock output includes sleep-wake, fasting-feeding, and energy expenditure cycles, while peripheral clock output includes a wide range of physiologic processes, including glucose homeostasis, oxidative metabolism, cytokine production, and stress response. The right column indicates different ways that circadian disruptors, such as diet, shift work, or other circadian rhythm sleep disorders, may impact the clock—i.e., by changing circadian period, phase, or amplitude. organismal rhythms in line with the external light/dark cycle. Under standing the circuit organization of the circadian clock within the brain is increasingly relevant in understanding how the master circadian pacemaker center within the SCN regulates feeding, sleep-wake activ ity, endocrine processes, energy expenditure, and metabolism (Fig. 498-2). Identification of the SCN as the master pacemaker was first established by the observation that SCN lesioning induced complete loss of rhythms of locomotor activity, drinking behavior, and endo crine hormone secretion. The ventral "core" region of the SCN, which is composed of neurons producing vasoactive intestinal polypeptide (VIP), receives photic information directly from the

retina through the RHT. At the molecular level, circadian gene transcription is induced within the SCN through the initial activation of immediate early genes, such as Per1, Per2, c-fos, and jun. Cells within the “core” region of the SCN then signal primarily via γ-aminobutyric acid (GABA)-ergic neu rotransmitter release to synchronize the cells within the “shell” region of the SCN, which produce arginine vasopressin (AVP), the most important neuropeptide for maintaining intra-SCN synchronicity. The SCN communicates to extra-SCN and peripheral clocks through both secreted factors and neuronal projections. The former was elegantly proven by the ability of SCN grafts to partially restore locomotor rhythms in SCN-lesioned animals. Efferent nerve outputs arise both from the AVP-producing shell region of the SCN and the VIP-predominated core. The SCN projects to several hypothalamic and thalamic regions, including the median preoptic nucleus (MPO), the subparaventricular zone (SPZ), the dorsomedial hypothalamus (DMH), the paraventricular nucleus of the hypothalamus (PVH), and the paraventricular nucleus of the thalamus (PVT). Some of these

Behavioral and physiologic outputs Circadian disruptors Sleep/wake Feeding/fasting Energy expenditure Glucose homeostasis Δ Period (High fat) Original phase New phase period amplitude Environmental light cycle Internal circadian time phase Δ Phase (Shift work) Δ Amplitude (Night eating, insulin resistance) Glucose homeostasis Lipogenesis Oxidative metabolism Mitochondrial respiration Xenobiotic detoxifixation Cytokine production Vascular tone Hemostasis Stress response Thermogenesis Incretin production DNA damage/repair CHAPTER 498 The Role of Circadian Biology in Health and Disease

regions, in turn, regulate output to both sleep- and wake-promoting regions, as well as to regions involved in regulation of autonomic, body temperature, and hormonal rhythms, as well as feeding. The SCN is thereby thought to promote sleep in part through the transmission of neural signals that terminate in the sleep-promoting ventrolateral preoptic nucleus (VLPO), i.e., one of the brain regions that is active during sleep. In contrast, the SCN promotes wakefulness during the active phase by transmission of neural signals that—by passing through regions such as the SPZ and the DMH—terminate in wake-promoting regions, including the locus coeruleus, lateral hypothalamic nucleus, ventral tegmental area, and dorsal raphe nucleus. The SCN also signals via noradrenergic fibers to the pineal gland to regulate the circadian production of the hormone melatonin. SCN con trol of the nighttime rise in pineal melatonin release (in both diurnal and nocturnal animals) is mediated through a pathway involving the PVH. Of note, artificial light at night delays the secretion of melatonin, ultimately affecting sleep (see “Endocrine Systems Regulated by the Circadian Clock” below). Melatonin plays a complex role in the circa dian system since the MT1 and MT2 melatonin receptors are expressed in the SCN itself; thus, melatonin feeds back to modulate circadian outputs to other cells in the brain and body. Neuronal output from the SCN also reaches peripheral tissues such as the adrenal glands, liver, and pancreas. The SCN produces rhythmic variation in multiple neuroendocrine axes, producing daily rhythms of gonadotropin, thyrotropin, and somatotropin. Prominent HPA axis rhythms ultimately give rise to daily variation in diverse pathways essential for hemodynamic stability, metabolism, and inflammation. These rhythms originate with SCN control of corticotropin-releasing

hormone (CRH)–producing cells in the PVH, which may regulate sleep as well as induce daily oscillations of both pituitary adrenocor ticotropic hormone (ACTH) and adrenal cortisol. Highlighting the importance of SCN output for peripheral rhythms, there is a dramatic reduction in the number of transcripts that exhibit circadian rhythms in the liver following SCN ablation in mice. Nonetheless, when the autonomous clock in the liver is ablated in mice, some key clock transcripts

such as Per2 still cycle provided the core body temperature rhythm persists. Whereas the SCN is exclusively entrained by light, meal timing can signal circadian time directly to peripheral tissues such as the liver. Thus, shifted meal timing as occurs during shift work or jetlag can uncouple peripheral clocks from the central pacemaker. Temperature can also phase shift peripheral tissue clocks, but not the SCN clock. This is an important phenomenon because, at the organis mal level, the SCN generates the core body temperature rhythm as one of the major mechanisms to signal circadian time to peripheral clocks.

■ ■ENTRAINMENT AND MEASUREMENT

OF THE CIRCADIAN SYSTEM Under normal light-dark cycles, the circadian system is corrected or “entrained” on a daily basis, producing diurnal rhythms of 24 h. Such signals of entrainment are called zeitgebers (German for “time-giver” signals) and include light exposure, meal timing, and activity patterns. Light serves as the dominant zeitgeber for the circadian system, and a breakthrough in understanding photoentrainment in mammals came with the discovery of the melanopsin system, which is composed of a specialized class of photosensitive retinal ganglion cells that expresses the blue light–sensitive photopigment melanopsin in the inner retina, separate from the photoreceptive rods and cones. Blue light around this wavelength (~480 nm) suppresses melatonin, such that melatonin levels are normally low during the day, promoting subjective and objec tive (electroencephalography assessed) wakefulness. PART 20 Emerging Topics in Clinical Medicine The ability of light to entrain the circadian system functions accord ing to a so-called phase response curve (PRC). When light exposure occurs prior to the critical phase of the core body temperature (CBT), defined by the CBT’s minimum, light produces a phase delay in the circadian rhythm. Conversely, light exposure after this critical period causes phase advances. The circadian system can respond even to small changes in light intensity (e.g., dim light at ~100 lux can produce half of the phase delay compared with an almost 100-fold greater light exposure). This responsiveness has been found to be highly individual and varies widely. This is in part due to genetic variation, as variants in clock genes can modulate the responsiveness of the human circadian system to light. When an organism is placed in an environment without zeitgebers, the circadian rhythm is said to free-run, as it relies on the endogenous rhythm of the circadian system. In humans, the study of endogenous circadian rhythms can be achieved by using a so-called constant rou tine that eliminates the risk of masking by factors such as sleep. In these paradigms, subjects are kept awake in a constant semi-recumbent posture, meals are provided on an hourly basis, and light is constantly kept below the level that can phase shift the SCN. Concurrently, circa dian rhythms are assessed by frequently measuring CBT, melatonin, or peptidergic hormone rhythms over the course of more than 24 h. In ani mals, endogenous circadian rhythms are instead studied by examining behavior, physiologic responses, and voluntary locomotor activity fol lowing 30–36 h of complete darkness. From these measurements, key properties of the circadian system can be ascertained, such as period length (peak-to-peak or trough-to-trough time), amplitude (peak-totrough difference), and phase (timing of peak or trough in relation to a reference point) (Fig. 498-2). These studies have revealed that the endogenous human circadian clock runs with a period length of ~24.2 h, while that of mice runs at ~23.5 h, with some variability across strains. In humans, evidence further indicates that females may have a slightly shorter circadian clock than males (24.1 vs 24.2 h), and many circadian parameters have been found to exhibit differences that are dependent on biological sex. Notably, interindividual variability in the endogenous circadian period length is further diversified by the existence of genetic polymorphisms in

clock genes (see below). These gene variants can confer extremes in the endogenous circadian period as well as phase; the latter can be advanced or delayed by ~3–4 h in each direction. This is due both to altered cir cadian rhythms at the cellular level and to altered SCN responsiveness to entrainment by light. For instance, PER3 gene contains a variablenumber, tandem-repeat polymorphism. Individuals homozygous for a PER3 5/5 genotype have been reported to be more responsive than PER3 4/4 homozygous individuals to the melatonin-suppressing effect of eve ning blue light exposure. By analyzing genetic variation across ~700,000 individuals, the number of genetic loci that have been identified that contribute to variability in chronotype is in the hundreds. Using specifically developed questionnaires to establish preferred sleep-wake timing, individuals can be categorized into so-called morningness-eveningness types or chronotypes. The most commonly used questionnaires are the Horne-Östberg Morningness-Eveningness Questionnaire (MEQ) and the Munich ChronoType Questionnaire (MCTQ). A composite MEQ score allows grouping into five categories that range from definite morning-type to evening-type individuals based on preferred waking time. In contrast, the MCTQ centers on the midpoint of sleep as a circadian marker, queries age and sex across a range of geographical locations, and can be used to ascertain differ ences between socially imposed sleep patterns (e.g., on working days) and sleep patterns on free days (the difference constituting so-called social jetlag). According to MCTQs obtained from primarily European populations, ~1% of the general population goes to bed before 10:00 p.m. and ~8% after 3:00 a.m. Differences in chronotype are linked to altered circadian timing, including peak levels of melatonin, which can vary by up to 4 h between extreme morning and evening types. Extreme chronotypes have also been shown to be linked to various traits; i.e., low morningness scores have been associated with greater tolerance to night shift work. Melatonin is one of the most commonly used peripheral markers of an individual’s circadian rhythm, reflecting the rhythmic function of the SCN. Circadian rhythms of melatonin can be measured in saliva or plasma, whereas 6-sulphatoxymelatonin (aMT6S), a metabolite generated from the breakdown of melatonin, can also be measured in urine. Accurate estimations of melatonin rhythms are often obtained by analyzing the dim light melatonin onset (DLMO). As the name implies, this involves evening/nighttime sampling of melatonin as opposed to 24-h sampling. This makes DLMO quantification useful in both the clinical and research settings. In normally entrained indi viduals, the DLMO can be used to ascertain whether an individual’s circadian rhythm is phase advanced or delayed, and this onset typically occurs ~2 h before the onset of sleep. The midpoint of sleep—the main marker used by the MCTQ—correlates more strongly with melatonin onset than the MEQ score. In the morning hours, the offset of mela tonin (“DLMoff”) can be used as a marker of circadian alignment or misalignment with the light-dark cycle. When individuals are exposed only to the natural light-dark cycle dictated by the sun, such as in an outdoor natural lighting environment, DLMO and DLMoff occur earlier. In contrast, exposure to artificial light has the overall effect of delaying the biological night and contributes to widening differences between chronotypes in modern society. The CBT is also often utilized as an indicator of the circadian rhythm. Even though CBT is more variable than DLMO, it usually correlates well with the phase obtained using the melatonin rhythm. The CBT, however, can be masked by factors such as sleep, food intake, and activity. CBT can be recorded and registered wirelessly with rela tive ease. In humans, CBT can be recorded via rectal thermometers or probes that are swallowed to pass through the gastrointestinal tract. When humans are studied under normal conditions with normal light ing and sleep duration from 2300 to 0700 h, the CBT reaches around 37.2°C by 0900 h, and from there, it continues to rise slowly until it reaches 37.4°C around 11 h later. The CBT then drops to the daily low of 36.5°C in the early morning (0400 h). The minimum in body tem perature also corresponds to the trough

in the 24-h rhythm in resting energy expenditure. Given the interrelationship between the circadian system and sleepwake systems, researchers have developed paradigms that uncouple

the circadian system from sleep-wake states, enabling the study of the contribution of the circadian system to investigated parameters across the entire sleep-wake cycle. These paradigms are known as “forced desynchrony” protocols and involve enforcing a significantly shortened (e.g., 20 h) or prolonged (e.g., 28 h) day length upon individuals. These protocols thus attempt to approximate what occurs during rotating shift work or “jetlag,” e.g., when travel across several time zones sud denly shifts the light-dark and behavioral cycles drastically away from the entrained 24-h rhythm. As described below, forced desynchrony protocols have contributed to uncovering how the circadian system regulates parameters such as cognitive performance, subjective alert ness, and metabolic and cardiovascular health. ■ ■PRIMARY PATHOLOGIES OF THE

CIRCADIAN SYSTEM An overarching term for disorders of the circadian system is circadian rhythm sleep disorders (CRSDs), where there is a mismatch between subjective behavioral and physiologic rhythms with the environmental light-dark or social activity-rest cycles (i.e., the body clock is out of sync with the external light-dark cycle). CRSDs can arise either due to misalignment of an exogenous environmental factor, such as light, or misalignment of the activity-rest cycle, such as occurs with shift work or jetlag, in relation to endogenous circadian timing. Mutations in the core clock genes themselves can also alter intrinsic circadian timing in relation to the external environment, which makes it difficult for these individuals to properly realign themselves. These disorders often result in adverse effects such as excessive sleepiness or depressed mood, often causing individuals to be unable to maintain a job or attend school at regular hours. The criteria for CRSDs based on the International Clas sification of Sleep Disorders (ICSD) are shown in Table 498-2. Animal models have greatly advanced our understanding of how core molecular clock components contribute to maintaining normal sleep-wake/rest-activity cycles (Table 498-3). For example, ClockΔ19/Δ19 mice have reduced total sleep duration and less induction of rapid eye movement (REM) sleep in response to sleep deprivation. Further, TABLE 498-3 Animal Models of Genetic Circadian Disruption AVERAGE CIRCADIAN TIME OF PEAK TRANSCRIPT LEVEL GENE MUTANT PHENOTYPE SCN PERIPHERY Bmal1 (Arntl) 15–21 22–2 Bmal1−/− Arrhythmic CK1δ (Csnk1δ) No rhythm No rhythm Csnk1δ+/– 0 to 0.5-h shorter period CK1ε (Csnk1ε) No rhythm No rhythm CK1εtau 4-h shorter period CK1ε — — CK1ε−/− 0.2- to 0.4-h longer period Clock No rhythm 21–3 Clock−/− 0.5-h shorter period — — — ClockD19/D19 4-h longer period/arrhythmic Clock/Npas2 — — Clock−/−/NPAS2−/− Arrhythmic Cry1 8–14 14–18 Cry1−/− 1-h shorter period Cry2 8–14 8–12 Cry2−/− 1-h longer period — — — Cry2A260T 0.2-h shorter period Dbp — — Dbp−/− 0.5-h shorter period Npas2 N/A 0–4 Npas2−/− 0.2-h shorter period Per1 4–8 10–16 Per1−/− 0.7-h shorter period — — — Per1brdm1 1-h shorter period — — — Per1ldc 0.5-h shorter period/arrhythmic Per2 6–12 14–18 Per2brdm1 1.5-h shorter period/arrhythmic — — — Per2ldc Arrhythmic Per3 4–9 10–14 Per3−/− 0 to 0.5-h shorter period Rev-erbα (Nr1d1) 2–6 4–10 Rev-erbα−/− 0.5-h shorter period/disrupted photic entrainment Rorα 6–10 Arrhythmic/various staggerer 0.5-h shorter period/disrupted photic entrainment Rorβ 4–8 18–22 Rorβ−/− 0.5-h longer period Rorγ N/A 16–20/various Rorγ−/− Normal behavior Note: Normal circadian rhythms of circadian clock and related genes, with description of circadian phenotype in mutant mice. Abbreviation: N/A, not applicable. Source: Adapted from Hum Mol Genet 15:R271, 2006, and Adv Genet 74:175, 2011.

TABLE 498-2 Criteria for Circadian Rhythm Sleep Disorders CRITERIA DESCRIPTION A A persistent or recurrent pattern of sleep disturbance due primarily to one of the following: • Alterations of the internal circadian timekeeping system. • Misalignment between endogenous circadian rhythms and exogenous factors that affect the timing or duration of sleep. B A circadian-related sleep disruption that leads to insomnia, excessive daytime sleepiness, or both. C A sleep disturbance that is associated with impairment of social, occupational, or other areas of functioning. mice that lack Bmal1 have increased total sleep time, but it is more fragmented and lacks clear 24-h sleep-wake rhythms, and mice lacking the repressors Cry1 and Cry2 are arrhythmic and spend more time in non-REM sleep. Finally, while ablation of the circadian gene Dbp does not alter the specific duration of sleep stages, it does lead to an altered circadian sleep-wake distribution, with more sleep during the normal wake period and vice versa. Consistent with a key role of clock genes in regulating sleep-wake behavior, human genetic studies of twins have found that up to half of the variation in diurnal preference is heritable. Established genetic variants associated with diurnal preference and cir cadian sleep disorders are listed in Table 498-4. The following briefly mentions the most common CRSDs, but readers should refer to the chapter on sleep disorders (Chap. 33) for a more detailed description. Delayed Sleep Phase Disorder Delayed sleep phase disorder (DSPD; or delayed sleep-wake phase disorder [DSWPD]) is one of the more common circadian rhythm sleep disorders, ranging from 0.2–16% of the population depending on definition used, and is most common in adolescence and early adulthood. DSPD is characterized by chronic and significant delays in both sleep onset and wake times compared to “socially acceptable” sleep-wake hours (i.e., “extreme night owls”). Rhythms of CBT and melatonin levels are also often
ALLELE

TABLE 498-4 Mutations and Gene Variants Linked to Sleep-Wake Disorders and Diurnal Preference GENE POSITION POPULATION SYNDROME/SLEEP PREFERENCE hCKIε S408N Japanese Protection against DSPS hCKIγ T44A Pedigree FASPS hCKIΔ H46R Pedigree FASPS hCLOCK T3111C (3′-UTR) European Eveningness hCRY2 A260T Pedigree FASPS hPER2 S662G (missense mutations in CKIε binding region) Pedigree FASPS hPER2 C111G (5′-UTR) British Extreme morningness hPER3 P415A/H417R Pedigree FASPS and seasonal affective disorder hPER3 G647 Swedish/Finish/Austrian/German Morningness hPER3 G647, P864, 4-repeat, T1037, R1158 Japanese DSPS hPER3 Increased repeats (exon 18, 54 bp) Brazilian DSPS hVIP rs9479402 (gene variant 54 kb upstream of VIP) European (>97% European ancestry) Morningness Abbreviations: DSPS, delayed sleep phase syndrome; FASPS, familial advanced sleep phase syndrome. delayed. DSPD is associated with polymorphisms within the circadian clock genes CLOCK, PER3, and CRY1, and the circadian period (tau) of these individuals may be longer. The most effective treatment includes bright-light therapy after waking in the morning (and/or dark-room therapy in the evening) in combination with melatonin administra tion in the evening several hours prior to the onset of sleep. These approaches attempt to realign endogenous circadian rhythms with the desired sleep-wake schedule, though often face challenges because individuals suffering from DSPD also phase delay more rapidly. Advanced Sleep Phase Disorder Another CRSD whereby one gets the correct amount and quality of sleep but at a shifted time is advanced sleep phase disorder (ASPD; or advanced sleep-wake phase disorder [ASWPD]). The prevalence of this disorder may be <1%, but the condition may be underreported, given that it may cause fewer conflicts with societal demands (i.e., 9-to-5 schedules) compared with DSPD. Individuals with ASPD experience an advance in their major sleep episode in relation to the desired

sleep-wake times. Thus, this disorder typically results both in very early evening bedtimes and morning awakenings (e.g., “extreme early birds”), resulting in reduced quality of life due to excessive sleepiness during the early evening, even in social situations. Individuals with ASPD also have phase-advanced temperature and melatonin rhythms. ASPD occurs more often in older individuals, although early-onset autosomal dominant familial variants (familial advanced sleep phase syndrome [FASPS]) have also been associated with mutations in either the PER2 or the casein kinase 1δ (CK1δ) gene. PER2 is critical for SCN resetting by light, and these PER2 mutations have been found to shorten the endogenous circadian period to ~23.3 h compared with the normal 24.2-h period length. Treatment includes bright light or blue-enriched phototherapy in the evening hours to delay the phase of the circadian clock to a later hour. Shift Work Sleep Disorder Given the increased prevalence of shift work in today’s 24/7 society and the accumulating evidence for increased incidence of sleep and metabolic disorders, including obesity, type 2 diabetes, cardiovascular disease, and cancer, in shift workers, the need to develop effective treatments for shift work sleep disorder (SWSD) is increasingly important. SWSD is at its core defined by the primary symptom of either insomnia or excessive sleepiness arising due to work scheduled during regular sleeping hours or at irregular times. The symptoms may arise because recovery sleep consumes a large proportion of the individual’s free time, potentially leading to negative social consequences such as difficulties maintaining social relationships. Older individuals are typically at an increased risk of SWSD due to the age-associated decline in the ability to maintain sleep during the time of day that would normally constitute the wake period. Therapeutic approaches include optimizing the sleep environment at home to minimize disruptions, melatonin prior to sleeping, and timed bright-light therapy. For example, for night workers, intermittent bright-light exposure during the night and avoidance of bright light

during the morning, even on days off, have been shown to improve sleep and feelings of alertness. Genetic screening combined with chronotype questionnaires may become useful tools for determining whether a given individual is suited for shiftwork. For instance, a twin study indicated that a genetic variant of the circadian gene DEC2 was associated with reduced sleep duration and shorter recovery sleep fol lowing extended sleep deprivation. More studies may reveal additional genetic variants that confer an advantage to repeated phase advances and phase delays as typically occurs in shift work. Irregular Sleep-Wake Rhythm Damage to the SCN can produce arrhythmicity in animals and is thought to be one of the possible underlying reasons for the temporally disorganized sleep-wake pattern that characterizes the disorder known as irregular sleep-wake rhythm (ISWR). Other contributing factors may be reduced responsiveness to entraining signals such as light and physical activity, as well as decreased exposure to such signals, as often occurs with increasing age. Despite normal total sleep time, there is a relative absence of a circadian pattern to the sleep-wake cycle such that sleep occurs in several distinct randomly distributed bouts. ISWR is often associated with neurologic impairment, foremost Alzheimer’s disease in older age; however, ISWR can also occur in individuals with poor sleep hygiene. Treatments involve multimodal interventions such as increased light exposure, improved sleep hygiene, and promotion of social and physi cal activities. Non-24-h Sleep-Wake Rhythm Disorder Individuals with non-24-h sleep-wake rhythm disorder (“non-24”), otherwise known as free-running disorder (FRD), have endogenous circadian rhythms that are not synchronized with the external 24-h day-night cycle due to an inability to readjust the circadian clock to the 24-h day on a daily basis. This most commonly occurs in individuals who are completely blind (i.e., lacking all photoreceptors) since they are unable to respond to daily light cues that normally would reset

the endogenous circa dian clock (although the condition has also been reported in sighted individuals). Instead, the sleep-wake period length corresponds to the individual's endogenous circadian rhythms, which are typically slightly longer than 24 h, thereby shifting sleep and wake cycles over time in relation to the light-dark cycle. Instead of sleeping at the same time each day, their sleep time would gradually be delayed each day until their sleep period literally goes "around the clock." Depending on the individual's endogenous rhythm, the individual will take a given number of days to realign their endogenous phase (in a 360° phase plot) with the zero time point in the exogenous 24-h light-dark cycle. Because of this chronic cycling, prominent symptoms of non-24 include sleep-wake cycle disruption (insomnia and daytime sleepi ness), impaired alertness and mood levels, and severe difficulties par taking in normally scheduled work, school, or social activities. Non-24 can be diagnosed following diurnal analysis of an individual's melato nin or cortisol rhythms, in combination with analyses of sleep diaries

where the sleep onset and offset can be visualized over time to identify the free-running period. Treatments for sighted non-24-h patients include a combination of bright-light therapy with appropriately timed melatonin administration, whereas melatonin and dual melatonin (MT1 and MT2) receptor agonist administration in completely blind non-24-h patients has been shown to entrain free-running rhythms and improve symptoms. Jetlag Most have experienced symptoms associated with jetlag, including insomnia, daytime sleepiness, and fatigue, when traveling from one time zone to another, as one's endogenous circadian rhythms are not yet aligned, or entrained, to the new external light-dark cycle. This is due to the slowness of the circadian system to adapt to the new time zone. Typically, the human circadian system can shift up to ~1.5 h a day in the westward direction (i.e., a phase delay), whereas it shifts more slowly (up to ~1 h daily) with eastward direction of travel (i.e., achieving a phase advance). Usually, symptoms of jetlag abate within the first couple of days after traveling and may present themselves after a first night of good sleep (which is more dependent on a high buildup of homeostatic sleep pressure). Older individuals (age >50) appear to be more at risk. While symptoms are transient, therapeutic approaches aim to hasten the synchronization of internal and external circadian cycles. Behavioral treatments include appropriately timed bright-light exposure and avoidance of bright light during the nighttime in the new destination, while pharmacologic approaches include timed melatonin administration before bedtime both prior to and following travel, resulting in improved sleep quality and decreased night waking. Social Jetlag Individuals with a late chronotype are prone to suffer from "social jetlag," a phenomenon in which individuals are forced to awaken at a point at which their bodies are entrained to be asleep due to discrepancy between alignment of social and biological time. Social SLEEP FASTING WAKE FEEDING CNS Inhibition of hunger Melatonin and GH secretion Neurotoxic substance clearance Memory consolidation Muscle Oxidative metabolism Adipose Lipid catabolism Leptin secretion Liver Gluconeogenesis Glycogenolysis Mitochondrial biogenesis Cholesterol synthesis Pancreas Glucagon secretion FIGURE 498-3 The circadian clock partitions behavioral, physiologic, and metabolic processes according to time of day. The partitioning of metabolic processes to appropriate times of day is critical for the maintenance of health from cellular to mammalian organisms. This figure highlights which processes peak within the central nervous system (CNS), muscle, adipose, liver, and pancreas during either the sleep/fasting or wake/feeding cycle in humans. GH, growth hormone.

jetlag can be estimated using questionnaires, such as the MCTQ, to compare sleep timing on working or school days compared with free days. This has established that a large proportion of the

European population suffers from 2 or more hours of social jetlag. Chronic social jetlag is associated with an increased risk of developing obesity and metabolic syndrome, as well as with greater alcohol consumption, smoking, and poorer academic performance in students.

The aforementioned categories of defined clinical circadian disor ders have been traditionally established based on consideration of the endogenous behavioral and physiologic cycles (primarily of melatonin and temperature) with the external 24-h light-dark cycle. In the fol lowing sections, we build on the concepts of circadian behavioral dis orders to consider new and emerging insight into the role of circadian disruption in organismal homeostasis (Figs. 498-3 and 498-4) and the availability of genetic strategies to dissect the interrelationship between clock function, health, and disease. ■ ■ROLE OF THE CLOCK SYSTEM IN PHYSIOLOGY Endocrine Systems Regulated by the Circadian Clock In addition to regulation of behavioral rhythms such as sleep-wake and fasting-feeding cycles, the circadian clock also regulates rhythms of the endocrine system. Cortisol rhythms are regulated through a feedback loop known as the HPA axis. Hypothalamic secretion of CRH and AVP promotes secretion of pituitary ACTH, which in turn regulates rhythmic cortisol secretion from the adrenal cortex. Cortisol release increases toward the morning, and this increase is believed to prepare the brain and peripheral tissues for daytime activity and food intake. AVP secretion in mice occurs prior to sleep to promote water intake, thereby preventing dehydration during the sleep period. Several hor mones, such as growth hormone (GH), cortisol, and melatonin, are influenced not only via circadian regulation, but also by sleep. For CHAPTER 498 The Role of Circadian Biology in Health and Disease
CNS Hunger signals Foraging behavior Cortisol secretion Neuronal activity Muscle Fatty acid uptake Glycolytic metabolism Adipose Lipogenesis Adiponectin production Liver Glycogen synthesis Bile acid synthesis Pancreas Insulin secretion

Circadian desynchrony CNS Depression Cognitive decline Pancreas Hypoinsulinemia Muscle Insulin resistance Sarcopenia Vasculature Adrenals Hematopoetic Chronic stress Disrupted HPA axis Autoimmunity FIGURE 498-4 Pathologies resulting from circadian desynchrony. Circadian rhythm sleep disorders, including advanced/delayed sleep phase disorder, jet lag, social jet lag, and shift work, result in a desynchrony between the environmental light-dark cycle “time” and the endogenous clock “time.” Pathologies can thus arise through misalignment imposed by exogenous (e.g., altered light cycle and/or feeding rhythm) and endogenous factors (e.g., mutations in core clock genes). Such desynchrony results in a host of wide-ranging pathologies across multiple tissues, including hypoinsulinemia (pancreas), disrupted hypothalamic-pituitary-adrenal (HPA) axis, autoimmunity, hypertension, obesity, and metabolic syndrome. CNS, central nervous system; IBD, inflammatory bowel disease. PART 20 Emerging Topics in Clinical Medicine instance, both GH secretion and the cortisol awakening response (CAR, i.e., the peak in cortisol soon after waking) are profoundly blunted by acute overnight wakefulness. GH secretion is primarily dependent on the occurrence of slow-wave sleep, which is a homeo statically driven sleep stage that occurs primarily in the first part of the sleep period. Both the CAR and daytime cortisol levels are also modulated by light exposure levels. Sleep also influences melatonin amplitude, such that sleep deprivation can increase melatonin levels. As sleep deprivation is often accompanied by artificial nighttime light exposure, the effect on melatonin can be combinatorial. In working environments, the effects of curtailed sleep are often confounded by mistimed exposure to light. The sensitivity to light levels that suppress melatonin can vary by an order of magnitude or more in individuals. This partly explains how even low levels of light can potently suppress melatonin secretion. Together with

altered timing in light exposure, perturbed hormonal levels likely represent a mechanism through which altered timing and duration of sleep can impact central and peripheral circadian oscillators. Centrally controlled rhythms of melatonin and cortisol are con sidered key regulators of extra-SCN and peripheral oscillators. Glu cocorticoid receptors exist in both the central nervous system and in peripheral tissues such as skeletal muscle, liver, and adipose tissue. Upon acute shifts in light-dark or feeding cycles, rhythmic levels in cortisol appear to modulate the rate at which behavioral and physi ological rhythms phase shift. Indeed, glucocorticoids regulate clock gene expression in muscle, kidney, and lung, and the powerful syn thetic glucocorticoid dexamethasone is often employed in vitro for its ability to synchronize (e.g., reset) circadian rhythms of cells, including liver cells. Consistent with a role for glucocorticoid regulation of the clock, both adrenalectomy, which results in a lack of cortisol, and exogenous corticosteroid supplementation significantly disrupt the circadian clock system. Several peripherally produced hormones and peptides are not only produced rhythmically but can also feed back to central clocks, includ ing within the SCN. For instance, both cortisol and thyroid hormones regulate their own rhythmic synthesis by feedback to central brain regions, i.e., the hypothalamus (for cortisol) and pituitary (for both hormones). Several other peripherally produced factors have been

Liver Circadian rhythm sleep disorders Dyslipidemia Steatosis Metabolic syndrome Jet lag Shiftwork Advanced/delayed sleep disorder Adipose Obesity Intestine Steatorrhea IBD flare Circadian dysbiosis Fibroblasts Thromboembolic events Hypertension Increased circulation of inflammatory cytokines Tumorigenesis proposed to influence the central clock, including fatty acids produced by the adipose tissue and fibroblast growth factor 21, a hormone pri marily produced by the liver. Peripheral hormones that signal energy state and hunger also exhibit circadian rhythms that are regulated by local tissue clocks. The most extensively studied of these hormones are leptin, which is released from white adipose tissue cells, and ghrelin, which is released from specific endocrine cells in the upper fundus region of the stomach. Ghrelin also exhibits significant peaks related to anticipated meal timing, which persist for several days of fasting in humans. Circulating rhythms of leptin and ghrelin are disrupted in circadian mutant mice and in humans experiencing circadian misalignment, with evidence for sex-specific effects. Per and Cry mutant mice exhibit severely blunted leptin rhythms, and wild-type mice exposed to jetlag (through repeatedly altered light-dark cycles) show a reduced wake-associated decrease in leptin. Similarly, humans forced to live 28-h days exhibit increased 24-h profiles of ghrelin and decreased levels of leptin. Ghrelin and leptin signal to several regions of the brain, including integrative appetitive regions of the hypothalamus such as the arcuate and paraventricular region. The response to these hormones is rhythmically regulated by the molecular clock within sev eral such central sites, effectively gating how these hormones influence rhythms of food intake and energy homeostasis in a time-of-day– and nutrient-dependent manner. Role for the Clock in Metabolic Homeostasis Circadian con trol of glucose homeostasis has long been recognized, as early studies demonstrated variation in glucose tolerance and insulin action across the day. For example, due to a combination of circadian control of both peripheral insulin sensitivity and pancreatic β-cell insulin secretion, oral glucose tolerance is lower in the evening and afternoon compared with the morning. Another example is the “dawn phenomenon,” whereby glucose levels peak prior to the onset of activity. Further, destruction of the SCN has been shown to abolish circadian regulation of glucose metabolism in rats, and daily cycles of insulin secretion and glucose tolerance are often perturbed in patients with type 2 diabetes, who also exhibit changes in gene expression rhythms in peripheral tissues such as adipose tissue. Changes in rhythmic parameters such as insulin secretion have also been observed

in first-degree relatives

of patients with type 2 diabetes, possibly highlighting a key hereditary role for the circadian clock in the pathogenesis of metabolic disease. Ablating clock genes in mice has revealed a key function for both central and peripheral clocks in regulating energy homeostasis. The circadian system has been shown to regulate rhythmic insulin secre tion from the pancreas via both neural signals and hormonal levels (e.g., cortisol and norepinephrine), as well as via cell-autonomous clock regulation within the pancreatic β cell itself. An early observa tion was that whole-body ClockΔ19/Δ19 mutant mice developed obesity without displaying hyperinsulinemia, a phenomenon that indicated concurrent β-cell failure. This was later confirmed using pancreas- and β-cell–specific Bmal1-deficient mice, which exhibited glucose intolerance, hypoinsulinemia, and impaired glucose-stimulated insulin secretion. The molecular clock within other peripheral tissues such as liver, adipose tissue, and skeletal muscle also regulates circadian fluc tuations in insulin sensitivity and glucose disposal, which are highest in the morning and decline toward the evening in humans. Liver-specific ablation of Bmal1 in mice has revealed that the liver clock promotes gluconeogenesis, glycogenolysis, and mitochondrial oxidative metabo lism in the sleep/fasting period, while promoting glycogen synthesis in the wake/feeding period. Muscle-specific Bmal1-deficient mice display reduced glucose tolerance, concomitant with lower levels of proteins involved in glucose uptake by muscle cells (e.g., the glucose transporter GLUT4). Ablation of the Cry1 and Cry2 repressors in the negative limb of the clock alters glucagon and glucocorticoid signaling in the liver, contributing to hyperglycemia and impaired glucose tolerance in these mutant mice. Together, these genetic studies in mice suggest a role for tissue-specific clocks in the partitioning of energy utilization across the sleep-wake cycle. Importantly, peripheral clocks also interact with other environ mental factors such as diet and time of feeding. For example, high-fat feeding leads not only to obesity and metabolic syndrome in mice, but also to perturbed clock gene expression across multiple peripheral tissues and a disrupted sleep-wake/fasting-feeding cycle, as revealed by increased activity and feeding during the daytime, the normal rest period in mice. Furthermore, mice that are fed a high-fat diet exclu sively during their (inactive) light phase gain significantly more weight than mice that are fed the same diet during the dark period—the active period for mice. Additionally, the metabolic phenotypes arising from ad lib high-fat feeding can be significantly ameliorated by restricting the time of high-fat feeding exclusively to the dark period. Animals with disrupted clock throughout the hypothalamus and SCN exhibit mistimed eating and adverse metabolic rhythms that can be restored by dark-only feeding. Time-restricted feeding can also increase the activity of brown adipose tissue in mice and reduce hepatic glucose production to instead promote beta oxidation of fatty acids. The poten tial clinical utility of time-restricted eating (TRE) has been corrobo rated in human interventional studies. These have demonstrated that dietary interventions modulate transcriptional rhythms across tissues and that TRE can improve metabolic homeostasis as well as promote weight loss. Compared with calorie restriction, TRE has repeatedly been shown to promote weight loss by reducing calorie intake without the need to actively count calories. Time-restricted eating may also modulate the central regulation of sleep and hunger, as studies have found that humans who restrict their food intake to a shorter than ad lib period also consume fewer daily calories and report both lower hunger and improved sleep. In the setting of critical illness, nutrition is often provided at the incorrect phase of the light-dark cycle, and interventions to align feeding with environmental zeitgeibers may improve metabolic health. There is also some evidence that consuming a greater proportion of daily calories early compared with later in the day confers metabolic advantages, including weight loss. One contrib

uting mechanism may be that diet-induced thermogenesis (energy expenditure elicited by food intake) is higher in the morning compared with evening and that daytime hunger ratings are lower when calories are preferentially consumed earlier versus later in the day. Finally, animal studies have further shown that when the light-dark cycle is disrupted or when animals are subjected to conditions mimick ing “jetlag” by artificially advancing or delaying the daily light period,

there is desynchronization among circadian clocks and subsequent weight gain. Accumulating evidence in humans also finds that circa dian misalignment both disrupts and desynchronizes circadian clocks across tissues. Clinical studies that have sampled tissues such as blood, skeletal muscle, and adipose tissue at regular intervals have observed disruptions in the day-night rhythms in clock and metabolic genes following sleep-wake interventions. Prolonged circadian misalignment using forced desynchrony protocols reduces insulin sensitivity in the pre- and postprandial states. Under such conditions, insulin secretion fails to suppress glucose levels, suggesting inadequate β-cell compensa tion. Moreover, resting metabolic rate declines significantly both in the awake and sleeping state, altogether providing potential explanations why shift work can increase the risk of obesity, type 2 diabetes, and the metabolic syndrome.

Human genetic association studies also support a role for clock genes in metabolic homeostasis and β-cell function. Carriers of a certain BMAL1 polymorphism have a greater risk of developing type 2 diabetes, while CLOCK variants have been found to interact with diet, such that variants can have a protective effect on insulin sensitivity in individuals with high monounsaturated fat intake or in individuals provided a low-fat diet. In contrast, the minor allele of another CLOCK gene variant has been associated with increased waist circumference, but only in those with high saturated fat intake. Similarly, NPAS2 and BMAL1 variants have been associated with a greater risk of hyper tension. Melatonin receptor MTNR1B gene variants that result in increased expression of MTNR1B have been associated with elevated fasting blood glucose levels and reduced insulin secretion irrespective of their level of glycemic control, although how melatonin regulates glucose homeostasis remains incompletely understood. These associa tion studies highlight the role of the circadian system in metabolism, as well as the potential for interactions of external perturbations—such as circadian misalignment—with a protective or adverse genetic profile. CHAPTER 498 A large proportion of society recurrently shifts sleep-wake and eat ing times between working/nonfree days and free days. This social jetlag has been increasingly tied to metabolic disruptions, including a greater risk of obesity and type 2 diabetes. As this involves recurrent phase advances and phase delays—like shift work but often of smaller magnitude—it is possible that social jetlag, and often interlinked eat ing jetlag, also results in perturbed rhythms of energy expenditure in combination with disruptions to the circadian hunger drive, further increasing the risk of obesity. Repeated shifts in the food- and SCNdriven rhythm of insulin release may similarly over time increase the risk of type 2 diabetes. Shifted feeding rhythms in relation to the sleepwake cycle and the timing of SCN activity may be causally involved in this pathogenesis. This is exemplified by the disorders known as nighteating syndrome and sleep-related eating disorder. In the former syn drome, a large part of daily calorie consumption occurs in the evening and nighttime hours, and this shifted meal pattern has been associated with a delayed timing of the internal clock. Some evidence exists that these syndromes are associated with obesity. Individuals who report sleeping fewer hours or who are subjected to restricted sleep for a few consecutive days have also been found to consume more calories, especially later in the evening, a period during which prolonged fast ing favors oxidative fuel utilization. As such, this may explain why

sleep restriction increases the risk of obesity. These associations have also been observed in individuals with later onset of sleep, i.e., evening chronotypes. Night-eating syndrome and later chronotypes have also been linked to type 2 diabetes and may be more common than other eating disorders such as binge-eating disorder. Both conditions have also been found to be associated with impaired glycemic control—such as a greater likelihood of hemoglobin A1c values exceeding 7%—in patients already suffering from type 2 diabetes. This emphasizes how proper alignment of internal circadian rhythms with external factors are key contributing factors for long-term metabolic homeostasis. The Role of Circadian Biology in Health and Disease
Circadian Clocks in Relation to Brain Health and Cogni tion Molecular circadian clocks are present not only within the extra-SCN regions of the brain, but also in neurons, astrocytes, microg lia, and cells of the blood-brain barrier. Emphasizing the functional

significance of properly aligned clocks for brain health, shift work ers have been found to have decreased gray matter in brain regions involved in memory and executive functions, with more notable effects in individuals who had shorter recovery periods between the onset of each shift work cycle. Adults performing rotating shift work for many years have also been shown to exhibit signs of accelerated cognitive aging. Notably, evidence suggests that these effects may be reversible, as those who stopped carrying out shift work exhibited normal cognitive performance 5 or more years later.

Studies have also uncovered an important role for perturbed circa dian and sleep-wake rhythms in neurodegenerative conditions such as Alzheimer’s disease (AD), Huntington’s disease (HD), and Parkinson’s disease (PD). Amyloid beta (Aβ), a key pathognomonic component of AD, normally exhibits circadian fluctuations in the extracellular space in the brain, as well as in the cerebrospinal fluid and plasma in humans, peaking during the active period and falling during sleep. Of note, these daily rhythms of Aβ accumulation are dampened in mice that are prone to develop AD; reduced plasma Aβ fluctuations have also been noted in older compared with younger individuals. Animal studies indicate that removal of Aβ (and other neurotoxic substances) during the nighttime sleep period is facilitated by a lymphatic-like system that relies on glial cells (the “glymphatic” system). Relevance of this system to humans is suggested by the observation that non–rapid eye movement (NREM) sleep is accompanied by hemodynamic fluc tuations that alter the flow of cerebrospinal fluid, which can remove toxins such as Aβ. Consistent with a role for circadian rhythms in the pathogenesis of AD, ablation of core clock genes throughout the brain, within subregions of the brain or within glia, leads to pathology such as oxidative stress, neuronal cell death, and scarring of brain tissue (astro gliosis). Furthermore, perturbed light-dark cycles increased pathology associated with oxidative stress, and single nucleotide polymorphisms in CLOCK and BMAL1 have been associated with increased risk of developing AD. PART 20 Emerging Topics in Clinical Medicine Evidence also indicates that the relationship between the circadian/ sleep-wake system and AD is bidirectional. For example, patients suf fering from AD exhibit several signs of perturbed circadian rhythms, the most prominent of such phenomena being “sundowning,” whereby AD patients become more agitated and exhibit delirium-like symptoms in the afternoon or evening. Studies have furthermore indicated that in severe forms of AD, the circadian rhythm is phase delayed. Aged AD-prone mice also display perturbed sleep-wake patterns, which can be corrected by immunization against Aβ or by an orexin antagonist. Further research will help uncover the primary role of the circadian system in disease pathology, independent of the contribution from perturbed sleep, in conditions like AD. Notably, evidence suggests that

interventions that increase daytime light exposure and that include melatonin supplementation ameliorate symptoms of AD, presumably by counteracting disrupted circadian rhythms. Meta-analyses of cohort and longitudinal studies support an asso ciation between shift work and the risk of depression, with greater risk in women. This relationship is bidirectional, as disruption of sleep and circadian rhythms is a key feature of depression and multiple other neuropsychiatric conditions. Two important factors for why mis aligned sleep increases the risk of neuropsychiatric conditions may be mistimed light exposure and disrupted 24-h rest-activity rhythms. In cross-sectional and longitudinal analyses, greater time spent outdoors during the day has been associated with fewer symptoms of insomnia, lower risk of developing depression, and less need for antidepressive medication. Similarly, decreased rest-activity rhythms are associated with lower subjective happiness and reaction time and a greater lifetime risk of major depressive or bipolar disorder. This may be partly due to genetic variation, as clock genes have been implicated in depression and mood both in human and genetic animal studies. Polymorphisms in genes that regulate sleep and circadian rhythms—for instance, a long gene variant of PER3—have also been linked to bipolar disorder and schizophrenia, while CRY2 and CLOCK gene polymorphisms are asso ciated with seasonal affective disorder, a type of depression arising in the fall and winter months when the levels of sunlight are lowest. Bipo lar disorder is furthermore often triggered by circadian disruptions

or curtailed sleep. Both bipolar disorder and schizophrenia have been linked to various forms of circadian disruption following disease onset, and a critical component of disease treatment often involves normal izing sleep and sleep-wake rhythms. Sleep deprivation by itself is known to reduce alertness, impair decision-making, and increase risk for accidents—after 18–24 h of con tinuous wakefulness, several skills exhibit the same degree of decline as following mild alcohol intoxication. However, cognitive abilities may suffer even further when sleep restriction is combined with circadian misalignment as in shift work. In one study, participants were subjected to ~43-h long days in parallel with reduced sleep (equivalent to 5.6 h of sleep in a 24-h period), yielding a forced desynchrony protocol coupled with sleep loss. When subjects were tested at the nadir of their circa dian period, the subjects' reaction speed dropped almost by an order of magnitude compared with controls. In another study, researchers noted almost a 36% greater incidence of serious medical errors in resident interns who regularly worked 24-h or longer shifts compared with those who were randomly assigned to work up to 16-h-long shifts. Furthermore, errors that resulted in patient death were three times more likely to occur in residents working extended hours compared with those who only worked up to 16-h-long shifts. Circadian Regulation of Gastrointestinal Homeostasis and the Microbiota Physiologic aspects of the gastrointestinal (GI) tract exhibit day-night variations that anticipate and prepare for food intake and digestion during the active period. Gastric emptying and colonic motility are considerably greater during the active phase, as the phasic motor program supporting movement of digested mate rial along the intestine is approximately twice as fast during the day compared with night. Bile acid secretion also exhibits circadian rhyth micity in the intestine, as do absorption and the expression of many nutrient uptake transporters in the intestinal wall, including the main glucose transporter protein SGLT1. The permeability of the intestinal wall also varies throughout the sleep-wake cycle, and mice exposed to chronic sleep fragmentation exhibit increased intestinal permeability, which may enable inflammatory molecules from bacteria to reach the systemic circulation. The composition and function of the fecal microbiome (i.e., the gut microbiota) also display circadian rhythmicity, orchestrated by both host circadian clock gene expression and food intake rhythms. Accord ingly,

circadian disruption, either by environmental or genetic means, perturbs these microbial rhythms, disrupting both bacterial levels and the metabolic functions of the gut microbiota. For example, alterations in the expression and functions of the gut microbiota have been noted in humans exposed to acute jetlag, and evidence suggests that curtail ing sleep, which often accompanies shift work and jetlag, can alter the gut microbiota. Corroborating the importance of the daily timing of food intake, interventions with meals scheduled earlier versus later in the day, or that involve time-restricted eating, have been found to alter the composition of the gut microbiome in humans, although the causal relevance of this remains to be ascertained. By increasing local and systemic inflammation, circadian disrup tion of the gut microbiota may be causally involved in increased risk of inflammatory bowel disease (Crohn's disease and ulcerative colitis) and colon cancer in shift workers. Biological sex differences have also been reported, as female mice display more pronounced microbial rhythms. Interestingly, the gut microbiome has also been shown to influence the rhythms of host tissues, such as the intestine and liver, that also appear sex-specific. This relationship indicates that a bidirectional interaction exists between tissues that regulate metabolic processes and the gut microbiome across the sleep-wake cycle. These findings may further more have clinical implications, given that the gut microbiome may both directly (in the gut lumen) and indirectly (through host-microbi ota interactions such as through signaling molecules) impact metabolic responses and pharmacokinetic and pharmacodynamic properties of therapeutic drugs across the 24-h day-night cycle. Cardiovascular Health and the Circadian Clock An early epidemiologic observation was a greater incidence of myocardial infarction in the morning hours, with the lowest risk during the

period preceding sleep. Other cardiovascular outcomes such as sudden cardiac death and syncope also exhibit a daily peak in the morning. Blood pressure (BP) typically peaks around 2100 h and decreases later during sleep. The postexercise recovery response of BP is faster in the late afternoon compared with the morning, and the daily timing of physical activity has been found to modulate the risk of all-cause and cardiovascular disease mortality. The lowering of BP during sleep is partially due to a circadian nighttime dip of around 3–6 mmHg in systolic BP (SBP) and 2–3 mmHg in diastolic BP (DBP). A dip in BP of either <10% or >20% during normal sleep has been associated with worse cardiovascular prognosis and risk of dementia. Nighttime BP dipping is also often disrupted in sleep-wake disorders and correlates with increased cardiovascular disease risk in conditions such as insom nia and narcolepsy. Conversely, specifically lowering nighttime BP has been found to confer a lower prospective risk of cardiovascular disease. In addition to BP, heart rate also typically decreases during sleep, while mistimed sleep leads to higher heart rate during the sleep period. Studies also suggest that heart muscle may be more tolerant to hypoxia and thus fare better under surgery scheduled for the afternoon due to timing of cellular programs driven by the cell autonomous clock in car diomyocytes. Thus, a combination of factors—which may also involve altered glucocorticoid levels and increased platelet aggregation—may contribute to a greater risk of cardiovascular disease in the morning. Subsequent epidemiologic studies also have demonstrated that shift work increases the risk of dyslipidemia and hypertension, as well as the risk of coronary heart disease, including myocardial infarction. These findings are in line with interventional findings in which circadian misalignment has been induced either by inverting the sleep-wake cycle or by imposing days that are far outside what the endogenous circadian clock can adapt to (i.e., either too short [e.g., 20-h] or too long [e.g., 28-h]). These studies in healthy human subjects have found that circadian misalignment elevates 24-h BP, particularly during sleep. These changes may be causally related to how the autonomic system is regulated during sleep, as

evidenced by reduced vagal cardiac control when the sleep-wake cycle is inverted. Circadian Disruption and Cancer In 2007, the International Agency for Research on Cancer (part of the World Health Organiza tion) declared that shift work that involves circadian disruption is likely carcinogenic to humans. While evidence for an association between shift work and general cancer incidence is mixed, accruing evidence supports a link between shift work and increased risk of developing colon and breast cancer, as well as having a poorer cancer prognosis. Telomere shortening, a phenomenon in aging that destabilizes the genome, has also been observed in shift workers as well as in individu als suffering from short sleep. Such changes may reduce the ability of damaged or senescent cells to undergo apoptosis and, instead, lead to uninhibited cell growth and cancer. An indirect role for the circadian clock has also come from retrospective studies on how cancer risk is related to food timing and duration of the nighttime fast in humans. In combination with interventional studies on time-restricted feeding, these findings suggest that limiting food intake to a restricted period of the day, optimizes circadian processes thereby reducing the risk of potentially carcinogenic cell damage. Studies of recurring fasting have also shown that it lowers the risk and delays the onset of cancer. Experimental genetic evidence has also implicated clock disruption as a factor in tumorigenesis. Genetic loss of Per2 or Bmal1 has been shown to promote lung tumorigenesis, while studies in Per2 mutant mice have also revealed increased radiation-induced lymphoma asso ciated with dysregulation of the cell cycle. However, disruption of the Cry gene in mice has also been implicated in tumor protection due to increased susceptibility to cell death. In contrast, pharmacologic overactivation of REV-ERB may impair growth of glioblastomas. While epidemiologic, experimental, and chronotherapeutic evidence (see sec tion “Chronotherapy and Future Directions”) suggests a link between circadian disruption and cancer, the precise role of circadian systems in tumorigenesis remains to be determined. Circadian Regulation of the Immune System Circadian mis alignment and sleep restriction both alter population levels of immune

cells and decrease the ability of immune cells to produce reactive radi cals, in part likely through disruption of cytokine rhythms. Chronic circadian disruption may thereby impair the immune system’s ability to conduct immunosurveillance at the proper time of day. This may reduce the ability to mount an appropriate pathogen-induced effector (cytotoxic T-cell) response during the active period, as well as impair the more long-term adaptive immune response, which is favored by the cytokine milieu (e.g., surges in prolactin and GH) that accompanies the recovery/sleep phase. Instead, circadian misalignment increases a range of clinically used inflammatory markers (e.g., C-reactive protein, tumor necrosis factor α, and interleukin 6), and such changes have been noted even when the sleep-wake cycle is only prolonged to a slightly longer than normal 24.6-h day. While similar effects are also observed following acute total sleep deprivation or recurrent partial sleep restriction, circadian misalignment has been found to promote an even more pronounced elevation of such markers. Genetic clock disruption in peritoneal macrophages has also revealed clock control of Toll-like receptor 9, which is responsible for identifying molecules from foreign pathogens. Clock knockout mice also have reduced T-cell antigen response, and mice immunized during the day had a stronger T-cell response than mice immunized at night, supporting regulation of the immune system by the clock. Similar mechanisms likely take place in humans, as clinical studies have noted an impaired vaccine response following sleep disruption, and several studies show improved immunogenic response to various antigens when vaccinated in the morning compared with afternoon.

Aging and the Circadian Clock Instability in the clock system is an often-overlooked hallmark of aging. Aging is associated with a decline in the robustness of intrinsic rhythmic processes at the behavioral, physiologic, and molecular levels in both human and animal models. At the behavioral level, aging leads to reduced and fragmented sleep, dampened locomotor activity and feeding rhythms, and a reduced ability to entrain to light, as old rodents are 20 times less sensitive to the entraining effects of light relative to young animals. Even middle-aged individuals exposed to jetlag exhibit more symp toms of circadian misalignment, such as increased time awake and reduced alertness, compared with young individuals. On a physiologic level, some of the hallmarks of aging are a reduction in amplitude (e.g., flattening of circadian pattern) of circadian processes, which can also be seen at the cellular level in peripheral cells isolated from older compared with younger individuals. This dampening of rhythms also impacts the circadian signal during the evening period (the wake maintenance zone). Epidemiologic evidence indicates that a dampened rest-activity amplitude is associated with an increased prospective risk of a range of common health conditions, such as dementia, CVD, can cer, and all-cause mortality. CHAPTER 498 The Role of Circadian Biology in Health and Disease

Aging also results in a phase advance (e.g., a shift in the timing of the peak or nadir) in rhythms of the endocrine and neuroendocrine sys tems, including sleep onset and offset. For example, cortisol, dehydro epiandrosterone (DHEA), and melatonin all have dampened rhythms and are phase advanced in aging; the combination of such changes may, for instance, contribute to more fragmented sleep and lower levels of restorative slow-wave sleep in aged individuals. Relatedly, aging results in reduced peptide expression in the SCN (VIP and AVP), cell loss in sleep-wake regions (including the SCN), and reduced amplitude of rhythms of SCN electrical activity. Further, while the SCN-dependent body temperature rhythm—a generally accepted marker for the integ rity of circadian rhythms—peaks in the evening and is lowest in the early morning in young individuals, aged healthy subjects display a phase advance and a decrease in circadian amplitude in body tem perature rhythms. Indeed, evidence suggests that internal desynchrony between core body temperature rhythms and the sleep-wake cycle may contribute to age-associated circadian alterations. On a molecular level, aging is associated with decreased expression and altered diurnal profiles of several of the core clock genes, includ ing Clock and Bmal1, within both SCN and peripheral tissues such as heart and liver. The acute induction of Per1 in response to light was markedly reduced in the SCN of aged mice compared with young mice,

potentially contributing to their delayed response to light entrainment. Mice lacking Bmal1 die prematurely compared with control mice, con sistent with premature accumulation of reactive oxygen species. These mice have an accelerated onset of numerous age-related pathologies, including cataracts, sarcopenia, reduced organ size, and decreased hair growth. Instead, deficiency of cryptochrome, a repressor of the core clock repressor, has been associated with alterations in liver regen eration, while BMAL1 and PER2 may be important for proper neuro genesis in the hippocampus, a brain region in which adult mammals normally exhibit continuous cell division. Altogether, this suggests that the highly conserved circadian clock is important for regulating a wide range of homeostatic processes, including cell-cycle pathways, which when properly phased to each other promote organismal fitness.

Shift workers have been found to exhibit molecular signs of accel erated aging, as measured by an accelerated DNA methylation clock. Measurements of altered circadian rhythms with age may serve as a useful biomarker for aging. An intriguing question is whether the decline in amplitude of

rhythms correlates with a decline in func tion and, importantly, whether restoration of these rhythms with age, through either behavioral or pharmacologic intervention, would delay the aging process. Studies in mice indicate that behavioral and pharmacologic interventions (including exercise) can restore circadian oscillations in aging. Restoration of levels of the metabolite NAD+, which are reduced with aging, in old mice by supplementation with the NAD+ precursor nicotinamide riboside (NR) markedly restores rhythms of metabolic and stress response pathways, as well as late evening activity rhythms, that decline with aging through inhibition of the clock repressor PER2. Similarly, transplantation of the SCN from a young rat into an old rat “rescued” the rhythms of both locomotor activity and corticotropin hormone (CRH), suggesting that the SCN is an important target for age-related changes in clocks. Physical activity or targeted therapeutics may therefore ameliorate some of the circadian deterioration in aged humans. PART 20 Emerging Topics in Clinical Medicine ■ ■CHRONOTHERAPY AND FUTURE DIRECTIONS Chronopharmacology, also known as chronotherapy or circadian medicine, is a rapidly emerging field that studies how the timing of drug administration may impact its effectiveness. Since physi ologic processes vary across the day, the timing of administration of medication may help optimize patient care. For example, since endogenous cholesterol synthesis is rhythmic in liver and peaks dur ing the early morning hours, administration of statins (HMG-CoA reductase inhibitors) in the evening prior to bedtime has proven to be more effective than daytime administration at reducing low-density lipoprotein cholesterol (LDL-C) levels because the highest concentra tion of medication coincides with the peak in rhythmic endogenous cholesterol production. Given that BP exhibits a 24-h rhythm—being lowest during sleep—angiotensin-converting enzyme (ACE) inhibi tors have been shown to be most effective at night to normalize the BP rhythms, restoring the nighttime dip in BP that is foremost tied to the occurrence of sleep. Numerous studies have also demonstrated that administration of cancer treatments at specific times of the day can increase chemotherapy effectiveness while also decreasing toxicity for a wide range of drugs. For example, 5-fluorouracil works best to treat colorectal cancer when administered at night, a time when the cancerous cells are more vulnerable while normal cells are quiescent and therefore less sensitive. Doxorubicin administration early in the morning to treat ovarian cancer has also been shown to be less toxic, as white blood cells recover faster than if the drug is given in the evening. Finally, the more severe morning symptoms of rheumatoid arthritis are linked to increased inflammation toward the evening; therefore, prevention of the nighttime upregulation of the immune/inflammatory reaction is more effective when glucocorticoids are administered with a nighttime release formulation. Recognition of circadian rhythms is also critical for diagnoses and treatment of endocrine disorders. The diagnosis of Cushing’s syn drome, which is characterized by hypercortisolemia, might be missed if the patient’s cortisol levels are measured in the morning, when endogenous cortisol production peaks. Therefore, clinical diagnosis

requires cortisol to be measured in the late evening when the levels of this hormone should typically be low. On the other hand, adrenal insufficiency is diagnosed by measuring cortisol in the morning when at its physiologic peak, and glucocorticoid therapy for these patients aims to mimic the endogenous rhythms of cortisol, as short-acting synthetic glucocorticoids are usually given several times a day in taper ing doses, such that the largest amount is taken in the morning and the smallest in the evening. Diabetes is another endocrine disorder intimately tied to circadian rhythms. Oral glucose tolerance, which is commonly used to diagnose diabetes, is worse in the afternoon and evening compared with the morning. This likely stems from greater daytime insulin sensitivity within peripheral tissues and reduced insulin secretion during the night. Similarly, due to

a surge in hormone levels in the morning, diabetes patients may suffer from the dawn phenomenon (or dawn effect), an abnormally high morning increase in blood glucose due to impaired response in insulin secretion. A related phenomenon that can be tied to evening timing of insulin doses is the “rebound” or Somogyi effect. In this scenario, the initially noted clinical sign in the form of elevated glucose levels may be noted in the morning. However, the underlying cause is hypoglycemia occurring during the night, which produces a counterregulatory hormonal response that subsequently results in morn ing hyperglycemia. As patients with type 2 diabetes often have grossly impaired daily cycles of insulin secretion and glucose tolerance, this further highlights that time of day is an important consideration for the diagnosis and treatment of metabolic disorders such as type 2 diabetes. Another example of potential clinical relevance is how the pharmacoki netics of metformin—the most common treatment for type 2 diabetes— is significantly impacted by time of day due to rhythmicity in glomerular filtration rate and renal plasma flow. Notably, large interindividual vari ability in the pharmacokinetics seems to stem mostly from differences in chronotype, highlighting the need for patient-specific treatments dictated by circadian gene-environment interactions. Continuous measurements of 24-h glucose have provided insight into sleep-wake regulation of glucose metabolism. Compared with day time glucose levels, nighttime blood glucose levels have been found to more accurately predict a range of glucoregulatory parameters. Emerg ing evidence has also indicated that the daily timing of exercise may be an important determinant for more efficacious improvements in blood triglyceride and glucose levels. Furthermore, consideration of meal timing, particularly in the hospital setting, may impact patient health or responsiveness to treatments, as food in hospitals is often provided either continuously or just during the dark (rest) phase, with the latter being common in neonatal intensive care. As our knowledge of the complexity of how circadian processes mod ulate physiology deepens, further advances to rationally develop new strategies for treatments of disorders affected by circadian misalignment are essential. For example, novel compounds have begun to emerge from unbiased drug discovery screens that in cell- and animal-based assays impact circadian clock components, either shortening or lengthening the period. These compounds include CRY stabilizers and various inhibitors of CKIδ, CKIε, and GSK-3. Pharmacologic control of the circadian cycle may be useful in the treatment of circadian disorders and metabolic disturbances with a circadian component. Understanding how the cir cadian clock controls biological functions will shed new light onto the pathogenesis of metabolic disorders with a circadian component, such as type 2 diabetes and metabolic syndrome, and will yield insight into how timing of drug delivery will impact patient care. Acknowledgment The authors would like to thank Billie Marcheva for her help with the figures and tables. ■ ■FURTHER READING Allada R, Bass J: Circadian mechanisms in medicine. N Engl J Med 384:550, 2021. Buxton OM et al: Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. Sci Transl Med 4:129ra43, 2012.

08 - 500 Emerging Neurotherapeutic Technologies

500 Emerging Neurotherapeutic Technologies

■ ■FURTHER READING Barabasi A-L et al: Network medicine: A network-based approach to human disease. Nat Rev Genet 12:56, 2011. Cheng F et al: Network-based approach to prediction and populationbased validation of in silico drug repurposing. Nat Commun 9:2691, 2018. Liu X et al: Robustness and lethality in multilayer biological networks. Nat Commun 11:6043, 2020. Loscalzo J et al (eds): Network Medicine: Complex Systems in Human Disease and Therapeutics. Cambridge, MA, Harvard University Press. Copyright 2017 by the President and Fellows of Harvard College. All rights reserved. Loscalzo J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007. Maiorino E, Loscalzo J: Phenomics and robust multiomics data for cardiovascular disease subtyping. Arterioscl Thromb Vasc Biol 43:1111, 2023. Menche J et al: Disease networks. Uncovering disease-disease rela tionships through the incomplete interactome. Science 347:1257601, 2015. Oldham WM et al: Network analysis to risk stratify patients with exer cise intolerance. Circ Res 122:864, 2018. Paci P et al: Gene co-expression in the interactome: Moving from correlation toward causation via an integrated approach to disease module discovery. NPJ Syst Biol App 7:3, 2021. Wang R et al: Multiomics network medicine approaches to precision medicine and therapeutics in cardiovascular diseases Arterioscl Thromb Vasc Biol 43:493, 2023. Jyoti Mishra, Karunesh Ganguly

Emerging Neurotherapeutic Technologies Neurotherapeutic technologies represent a diverse group of very promising treatment approaches with a common purpose of improving neurologic function. Decades of basic science research have paved the path for these novel technologies that have the potential to transform the lives of patients with neurologic diseases. A key goal is to minimize the consequences of lost abilities, whether they are motor, sensory, or cognitive. A common objective is to also harness the inherent plastic ity of the nervous system, regardless of age, and even in the face of a degenerative process. The technologies described below are the culmination of both an increased understanding of neural plasticity mechanisms in both the intact and the injured nervous system as well as advances in technol ogy and computational power. While it is also clear that

there may be fundamental limits on plasticity and repair mechanisms (the closing of developmental windows and/or loss of the ability of a network to com pensate), the brain remains highly plastic regardless of age and even in the face of ongoing injury and/or degenerative processes. Collectively, there is now growing evidence to support neurologic restorative efforts for both “static” (e.g., stroke) and progressive neurologic disorders. These technologies may not appear, at first glance, directly relevant to traditional medical care, but it is worth noting that clinicians have the most knowledge and experience about specific disease processes, available treatments, and the expected course of illnesses affecting the nervous system. It is thus critical that neurologic specialists and other clinicians play an important role in the future adoption of these technologies for neurologic rehabilitation. The sections below outline

emerging diagnostic and therapeutic approaches that have the poten tial to transform the lives of patients with neurologic disorders. These include technologies to harness plasticity, neuroimaging, neurostimu lation, and brain–machine interfaces.

NONINVASIVE TECHNOLOGIES TO HARNESS PLASTICITY Neurologic rehabilitation aims to harness activity-dependent plasticity mechanisms to maximize functional restoration. This principle can be applied to a diverse range of functional domains such as move ment control, sensory processing, language, pain, and cognition. For example, recent randomized controlled clinical trials for motor recov ery after stroke have suggested that intensity of training may be par ticularly important for sustained long-term improvements. Moreover, studies of the effects of such training in rodent and nonhuman primate models further suggest that plasticity of cortical “motor maps” as well as the coordinated firing of neurons in remaining networks underlie observed functional improvements with rehabilitation. The incorpora tion of technology for neurologic rehabilitation has the great potential to revolutionize the delivery of care by significantly increasing access, reducing the burden for adherence to high-intensity regimens, and maximizing engagement. Below are three examples of how emerging technology can be used to harness neural plasticity and maximize functional restoration. ■ ■ROBOTICS Rehabilitation robotics for both the upper and the lower limb have the potential to improve motor outcomes after stroke or other forms of brain injury. There is a growing recognition that focused training involving a range of tasks might be important for improved functional outcomes. While there is a growing recognition of “sensitive periods” that might represent optimal windows for rehabilitation after injury (e.g., perhaps the first several months after a stroke), such training likely has a role in the chronic period as well (e.g., maintenance therapy may also guard against known declines in function over time). Notably, the delivery of intensive training is a great challenge from both the per spective of the health care system and each patient. Outside of clinical trials, such a training program can be quite difficult to implement and maintain. It can also be costly and require significant effort. CHAPTER 500 Emerging Neurotherapeutic Technologies Motor rehabilitation protocols using robotics have been developed and tested for both the upper limb and the lower limb. Such robotic therapies have often focused on the delivery of high-intensity move ment practice that can surpass what is possible via existing standards of care. Moreover, robotic systems are capable of precisely measuring movement parameters (e.g., the kinematics of the movements) and providing quantitative feedback regarding the changes in performance during the training period. A particular focus has been on maximizing patient engagement and recruitment of attentional and reward path ways, both of which are increasingly recognized to drive neural plastic ity. Ongoing advances in design and the user interface will continue to improve comfort and support sustained effort. For example,

via close monitoring of performance and movement parameters, the system can aid at key points in order to minimize fatigue and ensure maximal engagement. Moreover, antigravity support of the upper limb can allow practice and task engagement even in the presence of severe weakness; this would be extremely challenging and labor intensive under current standards of care. Recent analysis also suggests that robotic devices may at least match outcomes realized with existing standards of care. However, rehabilitation robotics may also provide more precise feed back and permit novel quantitative rehabilitation approaches. Figure 500-1 shows one example of an upper-limb robotic exoskel eton device that is currently being evaluated for training after stroke. A randomized, multicenter trial compared treatment with this exo skeleton system against conventional therapy provided by physical and occupational therapists. Participants were enrolled in the chronic phase and all had moderate-to-severe deficits; the groups underwent three sessions per week over an 8-week period. For robotic training, subjects trained with games to improve mobilization and to practice activities of daily living. This study provided evidence that both conventional and

FIGURE 500-1 Photograph of a subject interacting with a complex upper-limb exoskeleton and a virtual reality system. (From U Keller et al: Robot-assisted arm assessments in spinal cord injured patients: A consideration of concept study. PLoS One 10:e0126948, 2015.) robotic therapy could improve function in patients with chronic stroke. Multiple studies have also found similar gains when using either con ventional or traditional approaches. Thus, a growing body of research supports the idea that such devices might complement conventional approaches to rehabilitation. Future work will need to define how reha bilitation robotics can optimally use adaptive and quantitative methods to further augment the recovery process. PART 20 Emerging Topics in Clinical Medicine ■ ■VIRTUAL AND AUGMENTED REALITY Therapeutic approaches using virtual reality (VR) and augmented reality (AR) aim to treat neurologic illnesses by specifically and quan titatively altering a patient’s subjective experiences and interactions with the environment. Core components of both are advanced hard ware and computational methods to generate simulated, yet realistic, perceptions. While some applications permit users to dynamically change the viewed perspective, other applications are designed to allow interactions among multiple users. Visual feedback is often a key com ponent; this can include simple computer monitors or more immersive “head-mounted” viewers that modify the simulation based on changes in perspective. Tracking of movements (e.g., hand and head position) is often included. Multiple methods are used to allow a user to interact with the environment; interactions can be guided by straightforward means such as a keyboard, mouse, or even a joystick. More immersive methods are also frequently used. For example, gloves with embedded sensors and haptic inputs can allow the user’s hand to be represented in real time in the simulated environment. Moreover, haptic interfaces can provide sensory feedback, allowing patients to interact with and “feel” virtual objects through multiple sensory modalities. A particular strength of these approaches is that therapeutic interventions can be studied in very controlled environments. VR enables a user to interact with a simulated reality that can be pre cisely and quantitatively controlled. In addition to allowing patients to dynamically experience an altered reality, it can simultaneously moni tor a subject’s behaviors and responses. Such monitoring can allow precise measurements of clinically relevant parameters (e.g., motor actions, perception, cognitive processing) and can also be applied in specific rehabilitation training to achieve functionally meaningful goals. A growing body of literature indicates that VR environments can be tailored to individual needs and preferences, thereby maximizing engagement, motivation, and adaptation to ensure sufficient difficulty of tasks. VR environments

can be designed to create powerful “gam ing” platforms that are actually targeting clinically relevant parameters. For example, the upper-limb robotic systems described previously are frequently combined with VR environments that allow interaction with virtual objects. In contrast to VR, AR overlays an artificial filter over a subject’s view of the actual physical world, thus providing an “augmented” or

enhanced view of the world around. AR is being tested in a diverse group of patients with neurologic impairments in the motor, sensory, or cognitive domains. AR may offer a particularly unique rehabilitation intervention for stroke patients. It is widely known that brain injuries limit patients’ physical interaction with their environments. Further more, physical and cognitive impairments may limit social interac tions. Such impoverished experiences are likely to be present during both the acute and the chronic phases. Importantly, there is clear basic scientific evidence that environmental enrichment can be a key com ponent of rehabilitation; such enrichment may offer additive benefits to the often-limited formal rehabilitation sessions per week. Consistent with this are clinical studies suggesting that motor and cognitive out comes may suffer when interactions with the environment are reduced; AR may be capable of increasing enrichment. For example, in the case of spatial neglect after stroke, the impaired modality may be accounted for using AR methods. Similarly, physical impairments that limit walk ing speeds can also limit visual feedback; both AR and VR can be used to enhance visual feedback during gait training. Figure 500-2 shows an innovative application of AR for the treat ment of “phantom limb” pain. A subset of both upper-limb and lowerlimb amputees experience painful sensations that appear to originate from the missing limb. Past research has suggested that mirror therapy can be an effective treatment for phantom limb pain. During mir ror therapy treatments, patients move their healthy arm in front of a mirror to produce a perception of movements of the missing limb. Previous studies have suggested that maladaptive plasticity of affected sensory cortices may be treated with mirror therapy. Importantly, in comparison to mirror therapy, AR-based therapy for phantom limb pain can be based on movements of the affected limb, i.e., using the remaining portion of the limb as opposed to the unaffected contra lateral limb. This study demonstrated a novel treatment in which “phantom motor execution” is enabled using sophisticated machinelearning algorithms. More specifically, the study “decoded” phantom limb movements by measuring electromyogram (EMG) activity at the stump. Importantly, while the distal muscles responsible for move ments were lost as a result of amputation, the remaining EMG activity could be used to predict presumed distal limb movements. As shown in Fig. 500-2, these inferred movements were projected onto an AR screen to create the perception of limb movements. The study showed that a subset of patients with long-term refractory phantom limb pain could experience a significant reduction in pain levels after using the AR system. ■ ■NEUROGAMING Computerized programs that harness the power of video games have shown some evidence for ameliorating deficits in visual perception, age-related degeneration, and neuropsychiatric disorders. An essential feature of effective video game training is the progressive adjustment of the level of difficulty in line with the cognitive improvement of the patient. Important areas of active research include ways to enhance sus tainability of neurogame training over long time periods and improv ing training transfer, i.e., the generalizability of task-specific training in one cognitive domain to more broad-based functional improvements. By leveraging video game technology, neurogames allow for dynamic user interaction and maintain user engagement across multiple ses sions over several days of training. Important game mechanics include repetitive practice, performance-adaptive challenges, and several lay ers of reward feedback—from moment-to-moment point rewards to reward milestones over multiple

sessions. Notably, neurogames have therapeutic potential as they can be targeted to specific neurocognitive deficits. For instance, games have shown significant benefits in aging, by targeting speed of processing and training the abilities to multitask and suppress distractions. In each case, selective targeting is achieved by focusing the adaptive chal lenges to the neurocognitive domain of interest. Duration of response time windows available to the user or the level of interference are selectively targeted in the case of speed of processing training and interference training, respectively. More recent research demonstrated that it is possible to engender focused circuit neuroplasticity using such

A B C D FIGURE 500-2 Augmented reality (AR) for phantom limb pain. A. A patient is shown a live AR video. B. Electromyography electrodes placed over the stump record muscle activation during training. C. The patient matches target postures during rehabilitation. D. Patient playing a game in which a car is controlled by “phantom movements.” (M Ortiz-Catalan et al: Phantom motor execution facilitated by machine learning and augmented reality as treatment for phantom limb pain: A single group, clinical trial in patients with chronic intractable phantom limb pain. Lancet 388:2885, 2016.) selective targeting in neurogaming. For example, older adults learned to adaptively perform within progressively more challenging distractor environments. Neuroplasticity selective to distractor processing was evidenced in this study at both the microscale, i.e., at the resolution of single neuron spiking in sensory cortex, as well as macroscale, i.e., elec troencephalography (EEG)-based event-related potential recordings. Video games have also shown promise in the treatment of visual deficits such as amblyopia, and in cognitive remediation in neuropsy chiatric disorders such as schizophrenia. However, while the evidence base has been encouraging in small-sample randomized controlled tri als (RCTs), larger RCTs are needed to demonstrate definitive therapeu tic benefit. This is especially necessary as the commercial brain training industry continues to make unsubstantiated claims of the benefits

 of neurogaming; such claims have been formally dismissed by the scientific community. Like any other pharmacologic or device-based therapy, neurogames need to be systematically validated in multiphase RCTs establishing neural target engagement and documenting cogni tive and behavioral outcomes in specific disorder populations. Generalizability of training benefits from task-specific cognitive outcomes to more broad-based functional improvements remains the holy grail of neurogaming. Next-generation neurogames will aim to integrate physiologic measures such as heart rate variability (an index of physical exertion), galvanic skin responses, and respiration rate (indices of stress response), and even EEG-based neural mea sures. The objectives of such multimodal biosensor integration are to enhance the “closed-loop mechanics” that drive game adaptation and

hence improve therapeutic outcomes and perhaps result in greater generalizabil ity. These complex, yet potentially more effective, neurogames of the future will need rigorous clinical study for demon stration of validity and efficacy.

NEUROIMAGING Feedback display (e.g., thermometer) ■ ■NEUROIMAGING OF CONNECTIVITY Multimodal neuroimaging methods including functional magnetic resonance imaging (fMRI), EEG, and magnetoen cephalography (MEG) are now being investigated as tools to study functional connectivity between brain regions, i.e., extent of correlated activity between brain regions of interest. Snapshots of func tional connectivity can be analyzed while an individual is engaged in

specific cog nitive tasks or during rest. Resting-state functional connectivity (rsFC) is espe cially attractive as a robust, task-indepen dent measure of brain function that can be evaluated in diverse neurologic and neuropsychiatric disorders. In fact, meth odologic research has shown that rs-fMRI can provide more reliable brain signals of energy consumption than specific task-based fMRI approaches. FIGURE 500-3 Neurofeedback using functional magnetic resonance imaging (fMRI). (From T Fovet et al: Translating neurocognitive models of auditory-visual hallucinations into therapy. Front Psychiatry 7:103, 2016.) PART 20 Emerging Topics in Clinical Medicine In recent years, there has been a surge of research to identify robust rsFC-based biomarkers for specific neurologic and neuropsychiatric disorders and thereby inform diagnoses and even predict specific treatment outcomes. For many such disorders, the network-level neurobiologic substrates that correspond to the clinical symptoms are not known. Furthermore, many are not unitary diseases, but rather heterogeneous syndromes composed of varied co-occurring symptoms. Hence, the quest for robust network biomarkers corresponding to com plex neuropsychologic disorders is challenging and still in its infancy; yet some studies have made significant headway in this domain. For example, in a large multisite cohort of ~1000 depressed patients, Drys dale et al. (2017) showed that rsFC measures can subdivide patients into four neurophysiologic “biotypes” with distinct patterns of dysfunctional connectivity in limbic and frontostriatal networks. These biotypes were associated with different clinical-symptom profiles (combinations of anhedonia, anxiety, insomnia, anergia, etc.) and had high (>80%) diagnostic sensitivity and specificity. Moreover, these biotypes could also predict responsiveness to transcranial magnetic stimulation (TMS) therapy. Another recent study demonstrated utility of rsFC measures to predict diagnosis of mild traumatic brain injury (mTBI), which is clini cally challenging by conventional means. Apart from fMRI-based measures of rsFC, EEG- and MEG-based rsFC measures are also being actively investigated, as these provide a relatively lower-cost alternative to fMRI. While EEG is of lowest cost, it compromises on spatial resolution. The major strength of MEG is its ability to provide more accurate source-space estimates of functional oscillatory coupling than EEG, as well as provide measures at various physiologically relevant frequencies (up to 50 Hz shown to be clinically useful). In this regard, EEG and MEG are complementary to fMRI, which can only be used to study slow activity fluctuations (i.e., <0.1 Hz); the potential for EEG/MEG modalities to provide valid diagnostic bio markers is currently underexploited and requires further study. ■ ■CLOSED-LOOP NEUROIMAGING Neuroscientific studies to date are predominantly designed as “openloop experiments,” interpreting the neurobiologic substrates of human behavior via correlation with simultaneously occurring neu ral activity. In recent years, advances in real-time signal processing have paved the way for “closed-loop neuroimaging,” wherein humans

3T MRI acquisition Image reconstruction The task of the subject is to lower the temperature display Real-time fMRI can directly manipulate experiment parameters in real-time based on specific brain signals (Fig. 500-3). Closed-loop imaging methods can not only advance our understanding of dynamic brain function but also have therapeutic potential. Humans can learn to modulate their neural dynamics in specific ways when they are able to perceive (i.e., see/hear) their brain signals in real-time using closed-loop neuroim aging-based neurofeedback. Early studies showed that such neuro feedback learning and resulting neuromodulation could be applied as therapy for patients suffering from chronic pain, motor rehabilitation in Parkinson’s and stroke patients, modulation of aberrant oscillatory activity in epilepsy, and improvement of cognitive abilities such as sustained attention in healthy individuals and patients with attentiondeficit hyperactivity disorder (ADHD). These approaches have also shown potential for deciphering state-of-consciousness in comatose

patients, wherein a proportion of vegetative/minimally conscious patients can communicate awareness via neuroimaging-based mental imagery. Closed-loop neuroimaging therapeutic studies have utilized realtime fMRI, EEG, and MEG methods. It is common for neural signals to be extracted from specific target brain regions for neuromodulation. However, given that distributed neural networks underlie behavioral deficits, new studies have also explored neurofeedback on combi natorial brain signals from multiple brain regions extracted using multivariate pattern analysis (MVPA). While early studies indicate therapeutic potential, clinical RCTs of closed-loop neuroimaging neu rofeedback have shown mixed results. This may largely be because of the individual heterogeneity in neuropsychiatric disorders such that there is no one-size-fits-all therapy. Closed-loop neuroimaging-based therapies need to be more personalized to the preintervention cogni tive and neurophysiologic states of the individual, and a better under standing developed regarding learning principles and mechanisms of self-regulation underlying neurofeedback. Clinical practitioners applying these methods also need to be well-educated on the hardware/ software capabilities of these brain–computer interfaces to maximize patient outcomes. NONINVASIVE BRAIN STIMULATION Noninvasive brain stimulation (NIBS) is widely recognized as having great potential to modulate brain networks in a range of neurologic and psychiatric diseases; it is currently approved by the U.S. Food and Drug

TMS coil Magnetic field TMS coil (μs) tDCS electrodes tDCS electrode Current flow – + + + + + + – – – – – – – – + + + (min) Anode Cathode FIGURE 500-4 Illustration of transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) setups. The upper panels show a TMS setup. Coils generate magnetic fields that can in turn generate electrical fields in the cortical tissue. The lower panels show a tDCS setup. The electrical current is believed to flow from the anode (+) to the cathode (–) through the superficial cortical areas leading to polarization. (Reproduced with permission from R Sparing, FM Mottaghy: Noninvasive brain stimulation with transcranial magnetic or direct current stimulation [TMS/tDCS]—From insights into human memory to therapy of its dysfunction. Methods 44:329, 2008.) Administration (FDA) as a treatment for depression. Importantly, there is a very large body of basic research indicating that neuromodulation of the nervous system with electrical stimulation can have both shortterm and long-term effects. While TMS uses magnetic fields to gener ate electrical currents, transcranial direct current stimulation (tDCS), in contrast, is based on direct stimulation using electrical currents applied at the scalp (Fig. 500-4). TMS induces small electrical currents in the brain by magnetic fields that pass through the skull; it is known to be painless and therefore widely used for NIBS. Animal research suggests that anodal tDCS causes a generalized reduction in resting membrane potential over large cortical areas, whereas cathodal stimu lation causes hyperpolarization. Prolonged stimulation with tDCS can cause an enduring change in cortical excitability under the stimulated regions. Further, changes in resting-state fMRI-based activity and functional connectivity have also been observed after tDCS. Notably, there is uncertainty regarding precisely how much electrical current is able to penetrate through the skull and modulate neural networks. Indeed, recent work has found that typical stimulation paradigms may not generate sufficient electrical fields to modulate neural activity; an alternate possibility is that peripheral nerves may be modulated and thus affect neural activity. Neuromodulation via stimulation techniques such as tDCS and TMS have shown promise as methods to improve motor function after stroke; there are a growing number of studies demonstrating functional benefits of combining physical therapy with brain stimulation. Two commonly utilized TMS paradigms include low-frequency “inhibi tory” stimulation of the healthy cortex or high-frequency “excitatory” stimulation

of the injured hemisphere. Each aims to modify the balance of reciprocal inhibition between the two hemispheres after stroke. A meta-analysis of RCTs published over the past decade found a significant beneficial effect on motor outcomes. Unfortunately, a recent large multicenter trial to assess the long-term benefits of TMS

on motor recovery after stroke (NICHE trial) did not find a benefit at the population level. Ongoing research aims to better understand how stimulation can directly affect neural patterns and thus allow more customization of stimulation—past trials did not record the neural responses to stimulation.

TMS and tDCS interventions are also being applied in psychiatric disorders. A substantial body of evidence supports the use of TMS as an antidepressant in major depressive disorder (MDD). TMS is also being investigated for its potential efficacy in posttraumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), and treatment of audi tory hallucinations in schizophrenia. Various repetitive TMS (rTMS) protocols have shown efficacy in major depression. These include both low-frequency (≤1 Hz) and high-frequency (10–20 Hz) rTMS stimula tion over the dorsolateral prefrontal cortex (DLPFC). Mechanistically, low-frequency rTMS is associated with decreased regional cerebral blood flow while high-frequency rTMS elicits increased blood flow, not only over the prefrontal region where the TMS is applied but also in associated basal ganglia and amygdala circuits. Notably, the differential mechanisms of low- versus high-frequency rTMS protocols are associ ated with mood improvements in different sets of MDD patients, and patients showing benefits with one protocol may even show worsening with the other, again pointing to individual heterogeneity in network function. EEG-guided TMS is also being investigated in psychiatric disorders, for instance, the individual resting alpha-band (8–12 Hz) peak frequency to determine TMS stimulation rates. With respect to transcranial electrical stimulation in psychiatry, tDCS is the most commonly used protocol. In major depression, there is a documented imbalance in left versus right DLPFC activity; hence, differential anodal versus cathodal tDCS in the left versus right prefrontal cortex may be a potentially efficacious approach. Interestingly while metaanalysis shows promise for NIBS methods in psychiatric illness, large RCTs have failed to generate benefits compared to placebo treatment. Future success may require careful personalized targeting based on network dynamics and refinement of protocols to accommodate com binatorial treatments. CHAPTER 500 Emerging Neurotherapeutic Technologies IMPLANTABLE NEURAL INTERFACES Fully implantable neural interfaces that can improve clinical function already exist. Cochlear implants, for example, are sensory prostheses that can restore hearing in deaf patients. Environmental sounds are processed in real-time and then converted into patterned stimulation delivered to the cochlear nerve. Importantly, even while the patterned stimulation remains the same, there are gradual improvements in the perception of speech and other complex sounds over a period of sev eral months after device implantation. Activity-dependent sculpting of neural circuits is hypothesized to underlie the observed perceptual improvements. Similarly, the development of deep-brain stimulation (DBS) was based on decades of work showing that surgical lesions to specific nuclei could alleviate tremor and bradykinesia in animal mod els. DBS involves chronic implantation of a stimulating electrode that targets specific neural structures (e.g., subthalamic nuclei or the globus pallidus in Parkinson’s disease). At least for movement disorders, it is commonly thought that targeted areas are functionally inhibited by the chronic electrical stimulation. ■ ■IMPLANTABLE DEVICES FOR NEUROMODULATION There has been recent progress in the development of implantable neu ral interfaces to treat neurologic and psychiatric illnesses. For example, for

patients with refractory focal epilepsy and clearly identified seizure foci, invasive “responsive stimulation” is FDA approved. Responsive stimulation is grounded on principles of closed-loop stimulation based on real-time monitoring of brain oscillations; specifically, the device aims to detect the earliest signatures of the onset of a seizure, usually at a stage that is not symptomatic, and then deliver focused electrical stimulation to prevent further progression and generalization. A large RCT of this device was performed in patients with intractable focal epilepsy; they were assigned to either sham or active stimulation in response to seizure detection. There was a significant reduction in

seizure frequency in the stimulation group, but it was rare for patients to become seizure-free. There were also modest improvements in quality of life. Nota bly, there was a small increased risk of hemorrhage associated with the device. In addition to providing clinicians with another treatment option, this device has offered important avenues for research and fur ther optimization. For example, it is now possible to monitor subclinical and clinical seizures and intracra nial EEG in patients with chronic epilepsy. This has resulted in new knowledge about the association of seizures with circadian rhythms and sleep. It is also anticipated that a better understanding of the triggers of seizures and the development of better stimulation algorithms, based on real-world data, can ultimately lead to more effective treatments.

Signal processing Neural signals

Action potentials Field potentials There is also great interest in the development of treatments for refractory depression. One area of focus has been on the development of DBS to treat depres sion. While early smaller studies were promising, a larger study failed to find benefits at the population level. Subsequent analysis has suggested the possibility that more precise tailoring of stimulation parameters to each individual is warranted, both at the level of specific pathways identified through neuroimaging as well as network activity biomarkers. Recent studies have, in fact, supported the notion that individualized patterns of net work activity are predictive of a patient’s symptoms and how the patient might respond to stimulation. There are now planned studies that aim to tailor stimulation to each individual with severe depression.

FIGURE 500-5 Components of a brain–machine interface (BMI). (Reproduced with permission from A Tsu et al: Cortical neuroprosthetics from a clinical perspective. Neurobiol Dis 83:154, 2015.) PART 20 Emerging Topics in Clinical Medicine ■ ■VAGUS NERVE STIMULATION TO IMPROVE RECOVERY AFTER STROKE Vagal nerve stimulation (VNS) has recently been approved by the FDA as a therapy to enhance motor recovery after stroke. Animal studies first provided clear evidence that VNS is safe and can enhance plastic ity in both intact animals as well as in models of injury. Importantly, these studies indicated that precise timing of movements is important for efficacy. For example, in animal models of stroke, stimulation of the vagus nerve was timed to the end of successful movement repeti tions; these studies further indicated that the precise timing of VNS during rehabilitation is essential. VNS appears to result in rapid activa tion of cholinergic and noradrenergic systems; the activation of these neuromodulators may enhance attentional effects and improve “signal to noise,” thus facilitating the encoding of relevant task features. This basic research culminated in smaller clinical trials and a subsequent pivotal randomized trial of VNS in stroke. In this trial, after 6 weeks of therapy paired with VNS, participants randomized to the VNS group (n = 53) had a significant increase in forelimb function compared to the control group. In

addition, 90 days after the study was completed, a higher percentage of patients in the VNS group maintained clinically meaningful responses. Together, this indicates that VNS is a promising new therapy to augment rehabilitation after stroke. However, given the variability of effects for single patients, additional research is required to determine which stroke patients are the most likely to benefit. Future advances that allow VNS to be delivered in the home setting should also lead to greater use of this approach. ■ ■BRAIN–COMPUTER INTERFACES FOR PARALYSIS Brain–computer interfaces (BCIs) represent a more advanced neural interface that aims to restore motor function. Multiple neurologic disor ders (e.g., traumatic and nontraumatic spinal cord injury, motor neuron disease, neuromuscular disorders, stroke) can result in severe and dev astating paralysis. Patients cannot perform simple activities, and they remain fully dependent for care. In patients with high cervical injuries, advanced amyotrophic lateral sclerosis (ALS), or brainstem strokes, the effects are especially devastating and often leave patients unable to com municate. While there has been extensive research into each disorder, clinically effective approaches for rehabilitation of long-term disability

Device control

Neural signals Control signals a Electrodes Computer cursor b Prosthetic limb Feedback are lacking. BCIs offer a promising means to restore function. In the patient groups described above, while the pathways for transmission of signals to muscles are disrupted, the brain itself is largely functional. Thus, BCIs can restore function by communicating directly with the brain. For example, in a “motor” BCI, a subject’s intention to move is translated in real time to control a device. As illustrated in Fig. 500-5, the components of a motor BCI include the following: (1) recordings of neural activity, (2) algorithms to transform the neural activity into control signals, (3) an external device driven by these control signals, and (4) feedback regarding the current state of the device. Many sources of neural signals can be used in a BCI. While EEG signals can be obtained noninvasively, other neural signals require invasive placement of electrodes. Three invasive sources of neural sig nals include electrocorticography (ECoG), action potentials or spikes, and local field potentials (LFPs). Spikes and LFPs are recorded with electrodes that penetrate the cortex. “Spikes” represent high-bandwidth signals (300–25,000 Hz) that are recorded from either single neu rons or multiple neurons (“multiunit”). LFPs are the low-frequency (~0.1–300 Hz) components. In contrast, ECoG is recorded from elec trodes that are placed on the cortical surface. ECoG signals may be viewed as an intermediate-resolution signal in comparison with spikes/ LFPs and EEG. While it is worth noting that there is still considerable ongoing research into the specific neural underpinning of each signal source, there has been great progress in the ability to decode a user’s intention. A central goal of the field of BCIs is to improve function in patients with severe disability. This can consist of a range of communication and assistive devices such as a computer cursor, keyboard control, wheelchair, or robotic limb. In the ideal scenario, the least invasive method of recording neural signals would allow the most complex level of control. Decades of research in nonhuman primates and early-phase clinical trials have demonstrated the feasibility of direct neural control of assistive technology based on recording of neural signals at multiple resolutions. There have been numerous examples of human subjects with a range of neurologic illnesses (e.g., brainstem stroke, ALS, spinal cord injury) who have demonstrated the actual use of implantable neu ral interfaces. This includes demonstrations of both the control of com munication interfaces as well as robotic limbs. Early pilot clinical trials of BCIs based on invasive recordings of neural signals showed that relatively high rates of brain-controlled typing are possible (e.g., >30

characters per minute). A past case study additionally demonstrated that a fully implantable BCI system could allow communication in a locked-in ALS patient (Fig. 500-6). At the time of the study, the patient required mechanical ventilation and could only communicate using eye movements. She was implanted with multiple subdural cortical electrodes; the neural signals were then processed and sent wirelessly to an external augmentative alternative communication (AAC) device.

A Posterior Anterior e1 e2 e3 e4 Electrode strip D Tablet Transmitter (implanted device) FIGURE 500-6 Illustration of an amyotrophic lateral sclerosis (ALS) patient with a fully implanted communication interface. A. Illustration of the location of electrodes on the brain. B. X-ray of chest showing the wireless module. C. X-ray of leads and wire routing. D. Schematic of the subject performing a typing task. (From MJ Vansteensel et al: Fully implanted brain–computer interface in a locked-in patient with ALS. N Engl J Med 375:2060, 2016. Copyright © 2016 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.) Importantly, she could use the interface with no supervision from research staff, albeit with a relatively low communication rate. Over the past 5 years, there has been tremendous progress toward the goal of restoring much higher rates of communication in partici pants with severe impairments. These studies have used either ECoG or spike-based decoding. One of the first studies indicated that a participant with a brainstem stroke and anarthria could communicate using a set of 50 words. Two subsequent studies showed that decoding a significantly larger set of words is possible, using either spiking or ECoG. For example, one study, using spike-based recordings, indicated that decoding of a large vocabulary was possible using phoneme-based decoding; that is, an arbitrary and a remarkably large set of words could be decoded by decomposing into its set of phonemes. Together, these studies indicate the real possibility of a clinically viable speech neuroprosthetic to restore fast communication in those with anarthria or severe dysarthria. Overall, there has been tremendous progress recently in the transla tion of BCIs. There are now also multiple commercial efforts to take these findings from pilot studies and to scale them to a commercially viable device. In fact, there is already a single participant with tetraple gia implanted with a first-in-class commercial device that can record spiking activity. While there are still challenges with long-term stability, this participant appears to be using this implanted device to control a computer (e.g., to control a cursor and to play video games) in the home setting. Additional work will be required to fully quantify how stable neural interfaces are and the level of performance that can be reliably achieved. As these characteristics become increasingly clear,

B C Electrodes (implanted) Ventilator Antenna CHAPTER 500 Receiver Emerging Neurotherapeutic Technologies it should allow targeted clinical translational efforts that are geared toward specific patient needs and preferences (e.g., extent of disability, medical condition, noninvasive vs invasive). For example, patients with high cervical injuries (i.e., above C4, where the arm and the hand are affected) have rehabilitation needs different from patients with lower cervical injuries (i.e., below C5–C6, where the primary deficits are the hand and fingers). Moreover, interfaces to restore communication may be different from those aimed toward movement control. We fully anticipate that over the next decade there will be larger scale clinical studies to demonstrate how BCIs allow participants with severe impair ments to experience the ability to communicate and to control assistive technology. ■ ■FURTHER READING Baniqued PDE et al: Brain-computer interface robotics for hand rehabilitation after stroke: A systematic review. J Neuroeng Rehabil 18:15, 2021. Bassett DS et al: Emerging frontiers of neuroengineering: A network science of brain connectivity.

Annu Rev Biomed Eng 19:327, 2017. Dawson J et al: Vagus nerve stimulation paired with rehabilitation for upper limb motor function after ischaemic stroke (VNS-REHAB): A randomised, blinded, pivotal, device trial. Lancet 397:1545, 2021. Drysdale AT et al: Resting-state connectivity biomarkers define neu rophysiological subtypes of depression. Nat Med 23:28, 2016. Ganguly K et al: Modulation of neural co-firing to enhance network transmission and improve motor function after stroke. Neuron 110:2363, 2022.

10 - 502 Metabolomics

502 Metabolomics

language processing relied on a specialized architecture called recurrent neural networks. Contemporary deep learning methods often leverage the transformer model (Table 501-2), which is well-suited to exploit the structure of natural language and other text. Text-processing machine learning models have been successfully applied to analyze physician notes in the electronic health record, detect depression symptom severity from spoken language, and scribe patient-physician visits. For example, a study by Rajkomar and colleagues analyzed electronic health record data from 216,221 adult patients to predict in-hospital mortality, 30-day unplanned readmission, and discharge diagnoses, among other outcomes, performing at high accuracy, with an AUC of 0.93–0.94 for predicting in-hospital mortality. Importantly, much of the progress in medical natural language processing has stemmed from the widespread availability of datasets, including, for example, the Medical Information Mart for Intensive Care (MIMIC) dataset.

Many specialized deep learning architectures have been developed for natural language processing applications, including the analysis of electronic health record data, using both supervised (e.g., recurrent neural network) and unsupervised (e.g., variational autoencoder) approaches. Domain-specific language representation models have been developed for the purpose of biomedical text mining, serving as a substrate for many downstream natural language processing tasks. Since ChatGPT was introduced in 2022, large language models including GPT-4 have rapidly been applied to diagnostic reasoning, health care documentation, and many other text-based tasks across medical specialties. The wide-ranging linguistic abilities and perfor mance of these models across myriad tasks have surprised many physi cians and machine learning practitioners alike. In a study by Kanjee and colleagues published in JAMA in 2023, the authors evaluated the general diagnostic reasoning abilities of GPT-4 on challenging medical cases published as part of the New England Journal of Medicine Clini copathological Conferences (CPCs), also known as the Case Records of the Massachusetts General Hospital. On these challenging cases, GPT-4, which was not trained specifically for medical diagnostic rea soning tasks, included the correct diagnosis as part of its differential diagnosis in 64% of the 70 cases assessed, a surprisingly high accuracy. PART 20 Emerging Topics in Clinical Medicine OTHER APPLICATIONS While medical computer vision and natural language processing tasks have been the focus of newer deep learning models due to the exten sive structure of imaging and text data, many other application classes exist. For example, cardiologist-level performance has been achieved in deep learning approaches for detecting arrhythmias from ambulatory electrocardiograms, standing in contrast to the rule-based algorithms used traditionally to interpret electrocardiographic signals. In genom ics, investigators have analyzed tumor genomes with machine learn ing methods to predict better survival using both deep learning and other machine learning approaches. Machine learning methods have also been used to characterize the deleteriousness of single nucleotide variants in DNA. Many other

applications of machine learning to new patient data streams are emerging, for example, machine learning applied to wearables (e.g., smartwatches). CONCLUSION Modern machine learning offers a powerful set of techniques to learn feature representations directly from data, already performing on par with expert physicians on select tasks. If carefully trained and judi ciously applied to key areas of clinician workflow, the representational power of new machine learning methods makes them likely to touch every area of clinical practice. ■ ■FURTHER READING Gulshan V et al: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402, 2016. Haug CJ, Drazen JM: Artificial intelligence and machine learning in clinical medicine. N Engl J Med 388:1201, 2023. Kanjee Z et al: Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 330:78, 2023.

Krizhevsky A et al: 2012 NeurIPS paper: Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012. LeCun YA et al: Deep learning. Nature 521:436, 2015. Olah C et al: Feature visualization. Distill. 2017. https://distill. pub/2017/feature-visualization/. Rajkomar A et al: Machine learning in medicine. N Engl J Med 380:1347, 2019. Ronneberger O et al: U-Net: Convolutional networks for biomedical image augmentation, in Medical Image Computing and ComputerAssisted Intervention – MICCAI 2015. Springer International Publish ing, 2015, pp. 234–241. Topol EJ: High-performance medicine: The convergence of human and artificial intelligence. Nat Med 25:44, 2019. Jared R. Mayers, Mathew G. Vander Heiden

Metabolomics Metabolism, loosely defined, represents the sum of all biochemical reactions involving small molecules with a molecular mass of ≤1000 Da within a given tissue, cell, or fluid. These small molecules are col lectively referred to as metabolites and are involved in the biochemical processes used to create macromolecules and fulfill the energy needs of a cell or organism. Metabolomics, then, represents the measurement of metabolites, either qualitatively or quantitatively, often as a way to gain insight into the metabolism of a cell, tissue, or organism. No one experi mental approach can characterize metabolism in its entirety; metabo lomics instead strives to measure a portion of the metabolome, which consists of all metabolites in a given biological sample at a given time. A link to a time-specific context is common to all “-omics” tech niques, but is particularly important in metabolomics. As metabolic processes are highly connected and interdependent, with individual metabolites often being involved in multiple pathways, levels of a specific metabolite can vary in response to an alteration in either the production or the consumption of that metabolite. Because significant changes in metabolite levels can occur over a very short time frame, the levels measured can be sensitive to perturbations either upstream or downstream of the measured metabolite in a pathway. This sensitivity can make measurement challenging, but it also makes metabolomics a powerful tool with which to assess either acute or chronic changes in cells or tissues. Indeed, the metabolome can be quite dynamic and reflective of the current condition of the material being assessed, as it ultimately represents an integration of outputs from the genome, epig enome, transcriptome, and proteome (Fig. 502-1). APPROACHES AND SAMPLING CONSIDERATIONS ■ ■UNTARGETED AND TARGETED METABOLOMICS There are two distinct approaches to measuring metabolites in biologi cal materials: untargeted and targeted metabolomics. These strategies differ in whether a predetermined subset of metabolites is intention ally sought in a sample, with the choice of approach dictated by the question under investigation. Regardless of the method utilized, it is important to recognize that no single metabolomics technique is comprehensive. Technical considerations heavily influence metabolite measurement, even with untargeted metabolomics, and no one method is able to capture the

entire metabolome. In this respect, metabolo mics contrasts with some other -omics techniques, like genomics or transcriptomics—i.e., in metabolomics, if something is not measured, its absence cannot necessarily be assumed.

Genome Epigenome Transcriptome Proteome Metabolome Phenotype FIGURE 502-1 The metabolome is downstream of the outputs measured by other “-omics” technologies. Thus, the state of the metabolome can more closely reflect clinical and experimental phenotypes. Untargeted Metabolomics Untargeted metabolomics is the com prehensive analysis of as many measurable analytes in a sample as possible, irrespective of their identity (Fig. 502-2). Among the benefits of this approach is that it is agnostic in its measurement of the metabo lome. Thus, it allows for the discovery of novel or unexpected mol ecules for further study. Coverage of the metabolome in an untargeted approach is influenced by the techniques used for sample preparation, metabolite separation prior to detection, and the inherent sensitivity and specificity of the analytical technique(s) employed (see “Metabolo mics Technologies,” below). A major drawback of untargeted metabolomics is that molecules of interest can be measured with less confidence or missed entirely because this approach carries an inherent bias toward the detection of high-abundance molecules. Handling and interpretation of data also represent a major challenge, as each sample run generates large amounts of data whose analysis can be both complicated and time consuming. Identifying each metabolite measured requires database searching, and further experimental investigation is often needed to confirm the exact identity of a signal of interest. Finally, in most cases, this technique yields only relative metabolite quantification, thereby rendering it most useful for comparisons between biological samples. Targeted Metabolomics Targeted metabolomics involves the measurement of a predefined group of chemically characterized metabolites—typically dictated by a hypothesis or predetermined plat form—with the aim of covering a select portion of the metabolome. The metabolites measured represent only a subset of those that would be measured by an untargeted approach; thus, a targeted approach generates a much smaller data set in which individual metabolites are detected with higher confidence (Fig. 502-2). Because the iden tity of each signal is known in advance, standards can be added to provide absolute quantification of each metabolite measured in the sample, although the use of targeted metabolomics to compare rela tive metabolite levels across samples is common. In addition, sample preparation and chromatographic separation before measurement can be optimized to improve detection of specific metabolites, enabling assessment of less abundant molecules. The key downside of targeted metabolomics is that information is gained about only those metabolites targeted by the analytical method. FIGURE 502-2 Untargeted metabolomics strives to measure as much of the metabolome as possible within a given biological sample, whereas targeted metabolomics focuses on measuring a predetermined subset of the metabolome. In untargeted metabolomics, a large number of signals corresponding to metabolites is generated, and further investigation is often necessary to assign a particular signal to a specific metabolite. Targeted metabolomics allows investigators to definitively measure signals that correspond to specific metabolites of interest.

■ ■SAMPLING CONSIDERATIONS Regardless of the approach used, it is important to consider potential sources of error that can influence the conclusions drawn from a metabolomic analysis. Because of the dynamic nature of the metabo lome, numerous biological confounders inherent to the samples them selves can affect levels of the metabolites measured. For this reason, the inclusion of controls or reference populations to account for these con founders can be critical for

data interpretation. Established biological confounders for patient-derived material include age, sex, body mass index, time of day collected, fasting status and/or dietary differences, and comorbid conditions such as diabetes or smoking. For example, metabolites commonly altered with respect to aging are those in anti oxidant and redox pathways as well as breakdown products of macro molecules. Sex differences influence a number of different metabolites, most prominently those involved in steroid and lipid metabolism. Per haps it is not surprising that diet can also affect the metabolome, and fasting has been shown to impact almost every category of metabolite frequently measured in biological fluids. Differences in sample handling and processing also influence metabolite measurements. Work using metabolomics to analyze mate rial from large prospective cohort studies has shown that changes in metabolite levels introduced by sample handling can lead to falsely positive associations between specific metabolite changes and disease risk. Specific considerations include the large geographic area of distri bution from which patients are drawn—e.g., a sample, such as blood, is collected locally and then exposed to variable conditions before being sent to a central lab for further processing. Moreover, because of the costs associated with obtaining and storing samples, often only one sample is available for each individual. CHAPTER 502 Metabolomics Time is a key variable in metabolite measurements, and efforts to assess the impact of sample handling and processing have led to improved analysis pipelines. For example, comparison of metabolites measured in samples undergoing immediate versus delayed processing can provide insight into those metabolites most affected by pre-pro cessing storage under varying conditions. More specifically, because metabolism occurs on a very rapid time scale, some metabolite levels will continue to change after sample collection even if the sample is stored on ice. Therefore, metabolism is ideally halted or “quenched” Untargeted metabolomics Targeted metabolomics

immediately via rapid freezing or chemical extraction, but practical considerations involved in the collection of material from patients can sometimes make rapid quenching impossible. Therefore, focus ing analysis on only those metabolites that are less sensitive to change due to delays in processing time may be important to gain biological insight.

Sequential metabolomic analyses of the same type of biological material from a patient can explore how metabolite levels vary over time. It is interesting that, when measured, many metabolites are found to be relatively stable. However, the extensive variability exhibited by some metabolites indicates that findings involving those metabolites should be interpreted with caution. Finally, the method of sample processing can affect which metabo lites are extracted from the material and thus influence what is measured. METABOLOMICS TECHNOLOGIES Metabolomics relies heavily on the intersection of instrumentation, software, and statistical and computational approaches for measure ment of metabolite levels and downstream data analysis. While the development of new and emerging techniques to assess the metabo lome is ongoing, the current, clinically applicable approaches can be separated into two broad categories: nuclear magnetic resonance (NMR)–based approaches and chromatography/mass spectrometry (MS)–based approaches. Each of these two approaches has its own set of advantages and disadvantages. ■ ■NUCLEAR MAGNETIC RESONANCE NMR is a technique that, at its core, exploits intrinsic magnetic proper ties of atomic nuclei to generate data. Nuclei with an odd total number of protons and neutrons (such as 1H, 13C, 15N, and 31P) have a non-zero spin, and this spin generates a magnetic field that can interact with externally applied electromagnetic fields. NMR places compounds into a magnetic field that induces the smaller magnetic fields to align with the larger one. Samples are

then exposed to a perpendicular elec tromagnetic field; the frequency of electromagnetic radiation needed to flip the spin of a nucleus in the exact opposite direction represents the frequency at which an atom “resonates” and can be measured. The resonance frequency of a given atom is affected by adjacent atoms and is ultimately unique for a given arrangement of atoms (i.e., each metabolite). This distribution or “spectrum” of signals is measured and recorded in an NMR experiment. PART 20 Emerging Topics in Clinical Medicine With respect to clinical applications, the primary benefits of NMR-based approaches are that they are nondestructive and can be performed on living samples, such as patients, cells, or tissues. They are also highly reproducible and require minimal sample preparation. Measurements are necessarily quantitative as the signal measured directly reflects concentration. These features ensure that multiple, comparable measurements can be made in a given sample either at a single point in time or across time. In addition, given that spins of different elements require sufficiently disparate resonance-inducing radiofrequencies in order to be entirely distinguishable, multiple ele ments can be assessed in a sample; this feature allows multidimensional Extraction Derivatization Chromatography Mass spectrometry data analysis FIGURE 502-3 Metabolite measurement by chromatography/mass spectrometry–based approaches involves multiple steps, and decisions made at each step influence what is measured. First, metabolites are extracted from a biological sample in a manner that is destructive of the original sample. This process stops biochemical activity and creates metabolite samples that can be analyzed, sometimes after a chemical derivatization step that alters a subset of metabolites in a manner that facilitates their downstream analysis. Second, metabolites in the sample are separated via chromatography. Finally, the chromatographically separated compounds are analyzed by mass spectrometry. Each signal detected corresponds to a metabolite’s characteristic mass per unit charge while the amplitude of that signal reflects the abundance.

TABLE 502-1 Comparison of Nuclear Magnetic Resonance (NMR)- Based and Mass Spectrometry (MS)-Based Approaches to Metabolomic Analyses FEATURE NMR MS Reproducibility High Lower Sensitivity Low (low µM) High (low nM) Selectivity Untargeted Targeted >> untargeted Sample preparation Minimal Complex Sample measurement Simple: single prep Multiple preps Metabolites per sample 50–200

> 1000 Identification Easy, using one- or twodimensional databases Complex; need standards and additional analyses Quantitation Inherently quantitative; intensity proportional to concentration Requires standards because of varying ionization efficiency Sample recovery Easy, nondestructive No Living samples Yes No cross-referencing of signals such as hydrogen and carbon. In an untargeted analysis, these multidimensional data can then be used for definitive metabolite identification, with comparison of results to known databases in which spectra for many metabolites in the human metabolome have been systematically recorded. Despite all these benefits, the primary challenge of NMR-based approaches is a lack of sensitivity. Because the time required to detect a signal is proportional to concentration, assessment of less abundant species is impossible or impractical. For example, while a typical NMRbased metabolomics analysis will return data on up to a couple of hun dred metabolites at concentrations of >1 µM, the MS-based approaches discussed below can

> distinguish more than 1000 metabolites at concen trations one to two orders of magnitude lower (Table 502-1). ■ ■CHROMATOGRAPHY/MASS SPECTROMETRY A distinguishing feature of chromatography/MS–based approaches is that a multistep process that destroys the material is necessary to generate a sample for analysis. In addition, each step of the sample preparation process involves decisions that influence the metabolites measured at the time of analysis. In general, once a sample to be ana lyzed is prepared, that material is subjected to a combined chemical and temporal separation of compounds via chromatography, with the output delivered to a device for performance of mass-based detection (technically, measurement of a mass-to-charge [m/z] ratio)—i.e., mass spectrometry. Finally, data collected by the mass spectrometer are analyzed (Fig. 502-3). Sample Preparation Although occasionally a part of NMR-based metabolite detection protocols, MS-based approaches almost uni formly require an initial sample-preparation phase called extraction. This technique destroys the original sample by partitioning metabolites into distinct immiscible phases, such as polar and nonpolar. These

phases are then mechanically separated and processed further for analysis. Given the nature of this extraction process, it is critical to determine in advance the general class of metabolites to be measured. This information will help to determine the optimal extraction pro tocol for specific types of metabolites of interest and to shape further downstream decisions regarding the chromatography/MS technique that also influences metabolite detection. In addition, depending on the metabolites to be analyzed and the method of separation and/ or analysis used, extracted samples sometimes are processed further in a preparative step called derivatization: extracted metabolites are chemically modified by the addition or substitution of distinct, known chemical moieties that facilitate separation or detection of types of metabolites. By changing the chemical properties of metabolites, derivatization may improve stability, solubility, or volatility or facilitate separation from closely related compounds, enhancing measurement of specific metabolites. Chromatography Chromatography is a ubiquitous approach used in chemistry for the separation of complex mixtures. The mixture of interest in a mobile phase is passed over a stationary phase such that compounds in the mixture interact with the stationary phase and transit through that stationary phase at different speeds, allowing their consequent separation. Two general types of chromatography are typi cally used in metabolomics. LIQUID CHROMATOGRAPHY Liquid chromatography–mass spec trometry (LC-MS) is the most commonly used approach in MS-based metabolomics. In this case, chromatography is characterized by a mobile phase that is a liquid and a stationary phase that is a solid. In liquid chromatography in particular, the choice of the solid and liquid phases can dramatically influence the types of compounds separated for input into the mass spectrometer. In general, LC-MS metabolomics is highly sensitive and versatile in allowing detection of a broad range of metabolites. A downside, however, is variability in exact separation timing, especially between different instruments; which metabolites are measured is impacted by the chromatography used and how well molecules are separated. GAS CHROMATOGRAPHY Gas chromatography–mass spectrometry (GC-MS) involves chromatography in which the mobile phase is a gas. In contrast to LC-MS, GC-based approaches have a narrower range of applications because only volatile metabolites that enter a gaseous phase are separated. When combined with

appropriate derivatization, GC-MS is a robust way to detect many organic acids, including amino acids, and molecules of low polarity, such as lipids. GC-MS is more reproducible than LC-MS across platforms and requires less expen sive instrumentation and less specialized training, but it also typically measures a much more restricted range of metabolites in a sample than does LC-MS. Mass Spectrometry Once the metabolites in a sample have been separated by chromatography, they are sent into the mass spectrometer for analysis and measurement. The first step in this stage of the process is to generate charged ions, as mass spectrometers measure compounds on the basis of their m/z ratio. Charge can be imparted through various techniques, although most commonly it is attained by either applying a high voltage to a sample or striking it with a laser. A number of different types of mass spectrometer can be employed for metabolomics. Three of the most commonly available types are discussed below. TANDEM MASS SPECTROMETRY Tandem MS relies on three sets of quadrupole magnets arranged in series. The power of this arrangement lies in its specificity through two sequential mass analyses of the same starting compound. In the first quadrupole, the “parent” or full ion is measured before being bombarded by an inert gas in the second quad rupole; this process fragments the compound into characteristic smaller “daughter” ions. The third quadrupole then measures these daughter ions. TIME-OF-FLIGHT MASS SPECTROMETRY While there are multiple types of time-of-flight (TOF) mass spectrometers, they all operate on similar principles. Most simply, lighter metabolites travel faster and

heavier metabolites travel more slowly. TOF machines have high mass accuracy and sensitivity while also acquiring data quickly.

ION TRAP MASS SPECTROMETRY Ion trap mass spectrometers, of which the orbital trap is a subtype, offer perhaps the highest degree of flexibility when it comes to MS-based metabolomics. In general, these machines can select for a specific mass range of metabolites at multiple levels, first by filtering with a single quadrupole and then by trapping and accumu lating metabolites of a particular mass or range of masses. This accumu lation can be applied to low-abundance compounds, allowing increased sensitivity. It also allows repeated fragmentation of metabolites (called MSn) to produce characteristic “daughter” ions, increasing the specificity of the analysis. Given this versatility coupled with high mass accuracy, the development of these machines is advancing rapidly; however, access to the latest versions can often be limited by cost. CURRENT CLINICAL APPLICATIONS Tests to assess small molecules are ubiquitous and well established throughout medicine. These include assays to measure select metabo lites of known clinical relevance, such as glucose, lactate, and ammonia. Of note, many standard tests assess these metabolites one at a time; however, metabolomics can allow the assessment of many metabolites in a sample and provide more information on metabolic state at a given point in time. In some cases, metabolomics is used to detect molecules for which there is not a robust single analyte test or when multiple species measured in a sample might provide new information. Here we will focus specifically on some applications of metabolomics tech niques in current clinical practice. CHAPTER 502 ■ ■MAGNETIC RESONANCE SPECTROSCOPY Magnetic resonance spectroscopy (MRS) is an adaptation of magnetic resonance imaging (MRI), a widely used technology in clinical prac tice. MRI, at its core, is essentially proton (1H) NMR with the result ing data rendered spatially to generate an image. Recall that NMR is nondestructive and can be applied to living samples. MRS, then, is a capability built into almost every MRI machine. In practice, radiolo gists can focus on specific volumes of interest within a patient’s imag ing and perform additional sequences to obtain an NMR spectrum in that space that can allow for the identification and quantification of specific

metabolites in that space. With this approach, several different metabolites across diverse classes, including lipids, sugars, and amino acids, can be measured at a given time. Metabolomics Extensive work has correlated different biological processes with altered levels and/or ratios of metabolites measured via MRS. One well-established application is in the diagnosis of brain masses. More specifically, N-acetylaspartate (NAA) is an amino acid derivative that is abundant in neurons, whereas choline is a metabolite whose level, as measured by MRS, correlates with cellularity and/or proliferation. Thus, an increase in the ratio of choline to NAA (and even loss of NAA signal entirely) correlates with cancer; tumors biologically are associ ated with the properties of increased cellularity from proliferation and the concurrent exclusion of normal neurons. A different process—for example, a brain abscess—does not result in increased choline levels (which instead may actually decrease), but does exclude neurons, resulting in an isolated NAA decrease. Metabolites such as lactate can also be helpful, depending on the clinical context, in providing insight into the metabolism of a tumor or identifying areas of early hypoxic brain injury after a stroke. Finally, among the several amino acids that can be measured, high levels of glutamine/glutamate can be helpful in a patient with altered mental status as changes in these amino acids are associated with hyperammonemia. (Glutamate serves as the cen tral nervous system sink for ammonia, generating glutamine in the process.) ■ ■NEWBORN SCREENING PROGRAMS Newborn screening programs are used to identify diseases within the first few days of life such that they can be treated or managed with early intervention. Among the classes of disease targeted by newborn screen ing programs are many inborn errors of metabolism, which often lead to changes in the levels of specific metabolites in blood or urine. One

of the first newborn screening programs tested for phenylketonuria, which results from the inability to metabolize phenylalanine resulting in high blood and urine levels of particular metabolites. Since that time, the panel used by programs throughout the United States and around the world has expanded dramatically. The general protocol is to collect a blood sample from infants in the first few days of life (often by heel prick on a piece of paper). These samples are sent to a central lab for analysis, which typically includes metabolomics measurements with targeted LC–tandem MS. Specific inborn errors of metabolism are sug gested by abnormal levels of a given metabolite or set of metabolites.

■ ■METABOLITE MEASUREMENTS IN CHILDREN AND ADULTS Outside the window of newborn screening, direct clinical measure ment of metabolite levels is also used in pediatric and adult patients. In these cases, clinical samples such as serum, cerebrospinal fluid, or urine are typically subjected to targeted LC–tandem MS to measure metabolites such as amino acids, acylcarnitines, and fatty acids. These measurements can help diagnose milder cases of inborn errors of metabolism that may have been missed by newborn screening. They can also help identify secondary metabolic defects, such as those that are related to nutritional deficiencies or are acquired in the setting of additional pathology. For example, these measurements are useful in determining the etiology of noncirrhotic hyperammonemia exposed by a catabolic stressor such as sepsis in a patient with a previously unknown subclinical or acquired urea-cycle defect. MS-based metabolomics is used by various athletic organizations for detection of metabolites associated with banned substances and by the pharmaceutical industry for assessment of levels of pharmaceuticals and their metabolites in both blood and tissues. Such analyses can provide key pharmacokinetic information to guide drug dosing and illuminate toxicology. These approaches can also be useful in clini cal practice. For example, chronic pain and its management remain a

challenge, and the sequelae of opiate/opioid use and abuse are of concern to many providers, their patients, and their patients' families. Therefore, many electronic medical records systems strive to ensure appropriate and consistent patient access to pain medications, while providers may need a means to ensure that patients are adhering to their prescribed regimens. One way to monitor drug use is to perform targeted LC–tandem MS for detection of specific drug metabolites in patients' urine. This approach is more sensitive than first-generation immunoassays and can detect a range of metabolites associated with other drugs beyond the one prescribed. Given that the first-generation immunoassays also often rely on confirmatory MS testing, upfront metabolomics reduces lab turnaround time and may also reduce costs by limiting multiple tests on the same sample. PART 20 Emerging Topics in Clinical Medicine EMERGING AND EXPERIMENTAL

CLINICAL APPLICATIONS The current clinical applications of metabolomics are largely limited to the indications described above. However, ongoing efforts are aimed at expanding the use of metabolomics for detection of biomarkers that can help with disease diagnosis or prognostication. ■ ■METABOLITES AS BIOMARKERS OF DISEASE There has been increasing work in prospective human cohort stud ies on the use of metabolomics, primarily MS-based approaches, to empirically identify small groups of metabolites whose altered levels are associated with the development or progression of disease. Efforts to characterize these "metabolic signatures" have been focused primar ily on common, multifactorial diseases such as diabetes, cardiovascular disease, and various cancers that are well represented in large prospec tive cohort studies. These studies have, for example, identified altered levels of amino acids that are associated with a future diagnosis of diabetes or pancreatic cancer. Similar efforts have proliferated across conditions ranging from chronic lung diseases to neurologic/develop mental disorders. Additional efforts have been made to assess the metabolome in patient samples at the time of an acute presentation. Because altered

metabolite levels can be associated with a specific clinical diagnosis and/or outcome, the idea is to identify a metabolite signature that facilitates diagnosis or provides prognostic information. This approach has been studied, for example, in the context of sepsis and septic shock, in which blood lactate levels are assessed in combination with the use of clinical tools such as the Acute Physiology and Chronic Health Evaluation (APACHE II) or the Sequential Organ Failure Assessment Score (SOFA). More recent efforts have identified a strong asso ciation between mortality and certain modified amino acids linked to mitochondrial dysfunction, highlighting a potential mechanistic link between sepsis pathogenesis and metabolic alterations. One key limitation in all of these studies is that researchers are pri marily assessing correlations between blood plasma metabolite levels and complex, multisystem diseases. It remains difficult to obtain a biological understanding of the mechanisms driving these changes or, even more simply, the primary tissue source(s) of these alterations from human data alone, without further experimentation in model systems. ■ ■REFINING DIAGNOSIS AND PREDICTION

OF DRUG SUSCEPTIBILITY In contrast to the above-described use of metabolomics-based approaches in multifactorial diseases, the application of these approaches in some specific contexts can yield an immediate diagnosis and suggest actionable therapeutic interventions. One specific example in oncology involves an understanding of the pathogenesis of oncogenic mutations in the metabolic enzyme isocitrate dehydrogenase (IDH) isoforms 1 and 2. The normal function of these enzymes is to interconvert isocitrate and α-ketoglutarate; however, cancer-specific point mutations in these enzymes alter the enzymes' function in a manner conferring

neomor phic activity that converts isocitrate into 2-hydroxyglutarate (2-HG). 2-HG is a metabolite that is typically present only at very low levels in cells, but when mutant IDH protein is present, 2-HG is produced and accumulates to high levels. Elevation of 2-HG can promote changes that directly contribute to malignancy; IDH mutations and 2-HG accumu lation are found in several human cancers, including specific clinical subsets of acute myeloid leukemia and glioma. Given the unique and specific accumulation of 2-HG in these mutant tumors, detection of this metabolite by LC-MS and NMR-based approaches has been studied both for diagnostic purposes and as a means of assessing drug response. For example, researchers have applied MRS-based approaches to assess the accumulation of 2-HG in gliomas, as this finding can noninvasively iden tify patients with an IDH-mutant subset of this cancer (Fig. 502-4). This diagnosis provides prognostic information and determines if a patient could benefit from drugs targeting mutant IDH that have been shown to benefit patients with IDH-mutant gliomas. In principle, metabolomics may identify other disease biomarkers to aid with diagnosis or therapy assessments in similar ways. ■ ■PHARMACOMETABOLOMICS The previous example positions metabolomics as a possible mecha nism for achieving a more personalized approach to medicine. The emerging field of pharmacometabolomics aims to take personaliza tion further by making this approach more widely applicable across drugs and disease states. The general idea is to link pharmacokinetics (PK) and pharmacodynamics (PD) data with baseline metabolomic profiling, with the goal of generating a predictive model for individual PK and PD responses based on a naïve patient’s metabolomic profile. Ideally, this approach would allow clinicians to take a baseline set of measurements and then—a priori—choose a specific dose of a specific drug to produce the desired effect in that specific patient. If successful, this method could limit both prolonged titration of medications and medication switching, dramatically shortening and simplifying the current approach to medical therapy. EMERGING TECHNIQUES While efforts to improve the existing capabilities discussed above are ongoing, innovations in instrumentation and computation are allowing collection and analysis of metabolite information that previously was not possible.

A B C FIGURE 502-4 In vivo 1H spectra and analysis demonstrating 2-hydroxyglutarate (2-HG) detection in isocitrate dehydrogenase (IDH)-mutant brain tumors. A–C. In vivo spectra from normal brain (A) and tumors (B–C) are shown. Components of 2-HG, γ-aminobutyric acid (GABA), glutamate, and glutamine are displayed. Measurement location is indicated by yellow box (voxel). 2-HG is seen only in mutant IDH brain tumors, but not normal brain or wild-type tumors. Shown in brackets is the estimated metabolite concentration (mM) ± standard deviation (s.d.). Cho, choline; Cr, creatine; Glu, glutamate; Gln, glutamine; Gly, glycine; Lac, lactate; Lip, lipids. Scale bars, 1 cm. (Reproduced with permission from C Choi et al: 2012.) ■ ■MASS SPECTROMETRY IMAGING Most clinical metabolomics relies on analysis of bulk material, but in an individual patient, there are areas of normal and diseased tis sue, and understanding the differences in metabolism in these areas requires both spatially sensitive resolution (imaging) and interrogation (metabolomics). While MRS can perform some of these functions, it is limited to macroscopic imaging (MRI) and relatively insensitive metabolomics approaches (NMR). In contrast, MS-based approaches, while more sensitive, by their nature rely on specimen destruction and homogenization. The premise of mass spectrometry imaging (MSI) is to overcome these limitations of MRS and mass spectrometry. MSI combines histologic evaluation of tissue with MS-based approaches to assess spatial differences in metabolites. MSI as a technique has been most highly refined in the neurosciences and can provide subcellular resolution. In general, thin slices of tissue are mounted on a slide, and metabolomics is performed at defined points across the slide, yielding spatial information on where in the tissue

section metabolites are mea sured. One specific approach utilizes matrix-assisted laser desorption/ ionization (MALDI) coupled to MS. In MALDI, tissues are coated with a special matrix and the MALDI laser scans point-by-point across a tis sue slice, ionizing the metabolites at each location for analysis by a mass spectrometer. These data can then be referenced back to an image of the original tissue slice (Fig. 502-5). This approach is being tested for defin ing tumor margins in real time during resection and thereby providing insight into boundaries between normal and abnormal tissues. ■ ■INTEGRATION WITH ADDITIONAL “-OMICS” TECHNIQUES There is increasing interest in integrating metabolomics data with data derived from other “-omics” techniques evaluating, for example, the Ionization FIGURE 502-5 Mass spectrometry imaging provides spatial information around metabolites in tissues. Tissue is mounted onto a slide, and a laser or another method is used to ionize metabolites in a discreet section of the tissue for detection by mass spectrometry. The process is repeated as the laser scans across the tissue, generating an “image” based on the levels of a metabolite detected at each point in the tissue section.

transcriptome or proteome (Fig. 502-1), an approach referred to as “multi-omics analysis.” Integrated multi-omics may provide a more complete understanding of the biological mechanisms underpinning observed phenotypes and is being used to study heterogeneity across cell populations determined via spatial and single-cell approaches. Additionally, when applied to complex communities like the gut microbiome, these approaches can aid in the discovery of previously uncharacterized metabolic pathways that impact human health. CHAPTER 502 ■ ■IMPROVING UNTARGETED METABOLOMICS Identifying unknown signals in an untargeted metabolomics analysis remains one of the central challenges in the field. As discussed above, NMR can definitively identify unknown signals but is inferior to MSbased approaches in its sensitivity and therefore in the number of signals it can detect in a given sample. To leverage the sensitivity of MS-based detection and overcome the challenge of metabolite identification, researchers are applying computational techniques, using networkstyle analyses and machine learning based approaches to streamline the process. The general approach is to combine information from known biological perturbations (e.g., changes in experimental conditions or disease states), empirical mass and structural information from MS analysis, and correlations with known metabolites/pathways to place unknown metabolites within existing metabolic networks. Metabolomics The growing interest in machine learning (a subset of artificial intel ligence), which focuses on the training of algorithms to analyze large amounts of data, has led to it being applied to facilitate analysis and interpretation of metabolomics data. These algorithms can be used to identify unknowns in untargeted metabolomics datasets and find pat terns of linked metabolites, or between metabolites and other data, that might otherwise be missed by traditional approaches.

12 - 504 Protein Folding Disorders

504 Protein Folding Disorders

Richard I. Morimoto, G. Scott Budinger

Protein Folding Disorders Many hundreds of human diseases, collectively known as protein con formational diseases or protein folding disorders, result from protein misfolding due to intrinsic and extrinsic errors amplified by aging and exposures to environmental and physiologic stress conditions. Such events challenge the functional integrity of the proteome and can lead to dysfunction, enhanced aggregation of proteins, mislocalization, and premature or inhibition of protein clearance, thus affecting cellular robustness, health, and longevity. Mismanagement of the proteome is therefore the basis of a broad class of diseases that includes orphan lysosomal storage diseases, type 2 diabetes mellitus, cystic fibrosis, some fibrotic diseases, metabolic diseases, muscle-wasting diseases, cancer, and neurodegenerative diseases exemplified by Alzheimer's disease, frontal temporal dementia, Parkinson's disease, amyotrophic lateral sclerosis (ALS), and Huntington's disease (Fig. 504-1). For each of these diseases and many others described in this textbook, aging is the major contributing risk factor. The challenge at the biochemical and molecular level is for the cell to achieve and maintain a stable and functional proteome during development that persists through young adulthood and is maintained throughout aging. For humans, this is essential for the operational health of each of the tens of trillions of cells that comprise our ~80 organs for health span and lifespan. To achieve this, our cells have evolved a remarkably efficient proteostasis network (PN) composed of ~3000 molecular chaperones and other highly conserved components essential for protein synthesis, folding, translocation, and degradation (Fig. 504-2) that balances input with output and ensures that every protein is functional. The PN is, therefore, essential for the robust ness of all tissues and for the diverse protein-protein interactions in cell signaling, biosynthetic processes, and the structural demands and mechanical requirements for tissue shape and function. An equal, if not more important, role for the PN is to detect, prevent, and remove PART 20 Emerging Topics in Clinical Medicine Eye Neuronal tissue Cataracts Corneal lactoferrin amyloidosis Hereditary lattice corneal dystrophy Lung Pulmonary alveolar proteinosis Cystic fibrosis Muscle Inclusion body myositis/myopathy Aortic medial amyloidosis Cardiac amyloidosis (e.g., transthyretin cardiomyopathy) FIGURE 504-1 Diseases of protein folding. A representative list of tissues affected and known folding diseases.

misfolded and aggregated proteins that accumulate in stress, aging, and disease and interfere with cellular health. Understanding how proteo stasis is achieved and maintained is of fundamental biological interest and essential to prevent age-associated protein folding disorders. Con sequently,

there is tremendous interest in the detection and treatment of these diseases, in particular as the human population continues to live longer. All organisms express an evolutionarily conserved set of molecular machines for the synthesis, folding, transport, and removal of unneces sary and damaged proteins. The PN is adapted for the highly specific physiologic requirements of tissues and the expression of abundant and rare proteins with wide-ranging solubilities, folding requirements, stability, and structural demands. Added to this complexity of natural clients for the PN is the additional load generated by genetic mutations carried in natural populations together with diverse environmental stressors that challenge PN capacity. Despite the central role of proteins as the workhorse of the cell, they are also highly susceptible to molecu lar damage, whether due to intrinsic metastability or to genetically inherited mutation or error-prone synthesis. Hence, dysfunction in the PN may clinically manifest as a gradual decline in homeostatic func tion in aging as occurs with genetic mutations or a loss of resilience in the face of environmental stressors. Thus, clinicians see the con sequences of proteostasis failure and cellular dysfunction in both the myriad disorders that present to physicians as age-associated clinical problems and the increased morbidity and mortality associated with trauma, infection, and other acute illness requiring hospitalization in older individuals. PROTEIN QUALITY CONTROL MECHANISMS The PN monitors and controls the flux of protein synthesis to promote functional folding and to minimize the accumulation of off-pathway aggregation-prone intermediates by their selective disaggregation or degradation. However, unlike an automobile assembly plant for which each part is designed and engineered for a specific use and precise assembly, the PN has built-in functional redundancy with properties to tolerate tremendous chemical noise and sequence diversity gener ated by coding region polymorphisms and biosynthetic errors among its client components. The PN has the ability to recognize and remove kinetically unstable conformational states of pro teins that accumulate in aging and that would otherwise compromise assembly and function. Proteins are highly sensi tive to fluctuations in their intracellular environment caused by shifts in ener getics, pH, and oxidizing and reducing conditions in addition to the myriad small molecules and metabolites that affect folding and function. Added to changes are the effects of external stress conditions caused by elevated tempera tures, infection, oxidizing and reducing environments, or osmolytes that can have profound consequences on pro tein folding thermodynamics, kinetics, and function. These intracellular and extracellular stress conditions, if not properly responded to, are predicted to further amplify protein instability from sequence polymorphisms and biosyn thetic errors that contribute to the stress of protein misfolding. The PN is orga nized at the cellular level into a series of highly coordinated molecular machines that direct the expression, biogenesis, and functional health of essentially all Alzheimer’s disease Amyotrophic lateral sclerosis Familial British dementia Familial Danish dementia Parkinson’s diease Huntington’s disease Thyroid Medullary thyroid carcinoma Immune system Systemic AL amyloidosis Multiple myeloma Pancreas/islet a cells Type 2 diabetes mellitus Liver α1 Antitrypsin deficiency Systemic diseases Lysozyme storage diseases p53-dependent cancers

Molecular chaperones Unfolded nascent polypeptide Native state Intermediate folded states Chaperones Autophagy Proteasome Normal turnover Misfolded states Degradation Toxic folds Improper trafficking Cystic Fibrosis Amyloidoses Emphysema Abeta, tau, Huntington, SOD1, α-synuclein α1 Antitrypsin Cystic fibrosis transmembrane conductance regulator FIGURE 504-2 The proteostasis network (PN) and folding diseases. Protein biogenesis through the action of molecular chaperones ensures the transition of the nascent polypeptide to on-pathway intermediates and the folded functional native state. Such proteins then have a normal turnover mostly through the

ubiquitin-proteasome system. Off-pathway species are prevented by the actions of chaperones and the recognition of nonnative misfolded states by the autophagiclysosomal pathway and the ubiquitin-proteasome system. When misfolded species escape quality control or overwhelm the PN, they can then become improperly trafficked as occurs for α1 antitrypsin associated with emphysema; for toxic folds as occurs for abeta, tau, Huntington, and SOD1 in amyloidogenic neurodegenerative diseases; and prematurely degraded as occurs for CFTR associated with cystic fibrosis. proteins (Fig. 504-2). More than to regulate and orchestrate these highly synchronized events, the PN is essential for protein quality control and for the prevention of the appearance of off-pathway con formational states and condensates, with accumulation of aggregates and amyloid species. Proteome health involves the constant exchange between the intrinsic physicochemical properties of polypeptides and the biological milieu of the cellular environment in which protein sequences and function evolved. Beginning with the synthesis of the nascent polypeptide on the ribosome, ribosome quality control (RQC) together with cytoplasmic chaperones of the HSP70, HSP90, DNAJ/HSP40, chaperonin/HSP60, and small HSPs (sHSPs) family ensure co- and posttranslational fold ing for the cell. Approximately 60% of the proteome resides in the cytoplasm and nucleus, for which the RQC, HSP70, HSP90, and HSP60 chaperones together with co-chaperones regulate the folded state of cli ent proteins through cycles of ATP binding and hydrolysis. Chaperones of the HSP70 and J-domain family are particularly well studied for their ability to interact transiently with short dispersed hydrophobic regions of nascent polypeptides using the energy from nucleotide hydrolysis to regulate the release of partially folded intermediates that either reenter the chaperone cycle or are released in a folded native state. For other chaperone clients, such as transcription factors, kinases, phosphatases, and signaling molecules, their folding to the functional state is highly regulated and dependent upon interactions with the HSP90 chaperone and other regulatory co-chaperones to form stable heteromeric complexes that hold the client in a partially folded state primed for subsequent regulated release. Consistent with the recognition that the formation of off-pathway aggregates is a kinetic component of proteostasis are the concerted activities of chaperone machines with disaggregase activity that unravel protein aggregates for refolding. These disaggregases include the AAA+ protein, ClpB in bacteria, Hsp104 in yeast and plants, and the function ally analogous metazoan disaggregase composed of HSP70, HSP110, and specific J-domain proteins. The subcellular organelles mitochondria and endoplasmic reticu lum (ER) account for ~20% of the proteome. Chaperone interactions are essential for mitochondrial-targeted proteins to maintain the extended polypeptide chain in a recognition-competent state for the

organellar receptors for translocation across membranes. Upon import, each translocated polypeptide is met by organellar-specific chaperones of the HSP70 and J-domain family for folding and assembly. While the mitochondrial genome encodes 13 proteins required for electron trans port, the great majority of mitochondrial proteins are encoded by the nuclear genome, synthesized in the cytosol, and imported across the outer and inner mitochondrial membrane. Hence, maintenance of the mitochondrial proteome relies on the coordinated efforts of both the cytosolic and mitochondrial PN. For translocation into the lumen of the ER, the extended polypeptide interacts with a set of glycosyl transferases, calnexins, calreticulins, disulphide isomerases, and lumen localized chaperones. Proteins that misfold in the ER are recognized and retro-translocated to the cytoplasm where they are directed to the ubiquitin-proteasome system (UPS) for unfolding, ubiquitination, and degradation.

The PN is balanced by the essential catabolic processes of the UPS and the autophagy-lysosomal pathway (ALP), which recognizes pro teins for degradation and recycling. The UPS is generally considered the primary pathway by which most proteins are recognized and tagged for degradation, and the ALP is highly responsive to nutrient deprivation and damage to recognize large aggregates and inclusions and engulf organelles and other subcellular compartments. In addition to their role in the regulated turnover of cellular proteins, these degra dation systems are essential for protein quality control and for limiting the accumulation of misfolded and aggregated proteins during stress conditions, aging, and disease. Protein turnover by the UPS involves an enzymatic cascade of E1, E2, and E3 enzymes that utilize ubiquitin and the recognition selectivity of E3s to tag clients, followed by degrada tion of the polyubiquitinated substrates by the 26S proteasome in the cytoplasm. Client specificity involves the large family of ~750 ubiquitin E3 ligases. In addition to their role of marking proteins for degradation, the ubiquitination machinery has numerous additional functions in cellular processes. For example, the ubiquitin ligase listerin is associ ated with the ribosome to ubiquitinate nascent chains that stall during protein synthesis to prevent the accumulation of aberrant polypep tides that would subsequently aggregate. Ubiquitination of nonnative aggregated clients by the ubiquitin ligase activity of the cochaperone carboxyl terminus Hsc70 interacting protein (CHIP) is central to the triage decision of the HSP70/HSP90 complex between client fold ing and proteasome-mediated degradation. ER-targeted clients that are misfolded are retro-translocated to the cytoplasm where they are polyubiquitinated and degraded by cytosolic proteasomes in a process termed ER-associated degradation (ERAD). Ubiquitination also pro vides crosstalk between the proteasome and autophagy pathways by targeting clients for lysosomal degradation and for endosomal sorting. Chaperones co-label a protein as damaged, recruiting other proteins that place ubiquitin chains on the damaged protein for degradation by the proteasome. Alternatively, chaperones can label proteins or protein aggregates to target them to the lysosome, and in this process, damaged proteins are degraded by the lysosome, an intracellular organelle with an acidic environment enriched in proteases through autophagy. CHAPTER 504 Protein Folding Disorders While there is a comprehensive understanding of the process of in vivo chaperone-dependent protein folding, the details of how these decisions are made for each client, whether and for how long to be maintained in a nonnative folding state through chaperone interactions in a nucleotide-independent state, or how to assemble into a stable chaperone complex for subsequent assembly into a functional state or to interact with chaperones to directly fold to a native state remain to be fully addressed. CELL STRESS RESPONSES: SENSORS AND REGULATORS OF PROTEIN DAMAGE Cell stress responses are ancient genetic networks that detect, adapt, and protect all cells against toxic environmental stressors and physiologi cally relevant changes in their cellular environment, including changes induced during development and tissue repair after injury (Fig. 504-3). At the core of these cell stress responses are molecular switches: (1) the heat shock response (HSR) that protects proteins in the cytoplasm and nucleus regulated by HSF-1; (2) the unfolded protein response (UPR)

Stress responses Programmed Repression of the Heat Shock Response, UPR, and Oxidative Stress Response Molecular chaperones Protein quality control Proteostasis Disease Aging Reproduction Development High Risk Low Risk FIGURE 504-3 Aging and protein folding diseases. Aging is the major risk factor for degenerative diseases. Cell stress responses (heat shock response and the unfolded protein responses [UPR] in the endoplasmic reticulum and mitochondria) decline at reproductive maturity in studies from Caenorhabditis elegans and prevent adaptive and protective increased expression of molecular chaperones to prevent protein misfolding. of the ER (UPRER)

controlled by XBP1, ATF6, and ATF4; (3) the UPR of the mitochondria (UPRmt) controlled by ATFS1; (4) the DAF-16/ FOX-O stress response pathway associated with insulin signaling; (5) the integrated stress response (ISR) controlled by PERK, PKR, HRI, GCN2, and ATF4; and (6) the antioxidant stress response regulated by NRF-2. Collectively, these cell stress responses and their respective transcription factors (TFs) are essential for all cells and tissues regu lated both autonomously and cell nonautonomously across tissues in metazoans to detect proteotoxic stress, to adapt and protect the cell against the toxic consequences of protein damage, and to regulate changes in the proteome necessary for differentiation. While each of these cell stress pathways can be activated independently, they are also induced in different combinations according to the chemical and physiologic properties of the stress signal(s) and provide crossprotective mechanisms. PART 20 Emerging Topics in Clinical Medicine The HSR is an evolutionarily conserved cellular defense mech anism that protects cells against proteotoxicity associated with misfolding, aggregation, and proteome mismanagement. HSF-1 inducibly regulates the transcription of genes encoding molecular chaperones and components of the PN. In unstressed cells, HSF-1 exists in an inert monomeric state in the cytoplasm or nucleus where it is negatively regulated by the molecular chaperones, HSP70 and HJD-1. Upon exposure to heat shock, HSF-1 undergoes a series of molecular transformations and rapidly trimerizes to acquire DNAbinding activity, undergoes extensive posttranslational modifications by phosphorylation and sumoylation, binds to heat shock elements in promoters of heat shock responsive genes, and associated with these events, forms nuclear stress bodies. Upon dissipation of the stress signal, the HSR attenuates by the active repression of HSF-1 DNA binding by acetylation and loss of HSF-1 transcriptional activity. This is accomplished by binding of HSF-1 with HSP70, HDJ-1, and HSBP1, leading to its dissociation from the trimer to the monomeric state. In addition to HSF-1 being essential for the HSR and cell and organismal stress resilience, HSF-1 is essential during early develop ment in metazoans, functions as a maternal factor for gametogenesis, regulates oocyte maturation by activating genes that function in the meiotic cell cycle, is constitutively activated in cancer, and is neces sary to maintain NAD+ and ATP levels. In the ER, the UPRER involves three stress response arms regulated by the transcription factors XBP1, ATF6, and ATF4 that bind to specific cis- elements for these ER-stress-responsive pathways. XBP1 is activated by IRE1, which is a transmembrane protein with kinase and endoribonu clease (RNase) activity that senses misfolding in the ER directly, leading to its autophosphorylation, oligomerization, and acquisition of RNase activity. This allows active IRE1 to cleave XBP1 mRNA, generating a spliced transcript (XBP1s) that encodes XBP1 to induce the transcription

of UPR target genes. ER stress also promotes the relocalization of ATF6 from the ER membrane to the Golgi apparatus, where it is cleaved by the proteases SP1 and SP2, generating a cytosolic fragment of ATF6 that translocates to the nucleus to direct transcription of a complementary set of UPR genes. Together, XBP1 and ATF6 induce the expression of genes involved in protein folding, ER-associated protein degradation, and lipid metabolism. A third ER transmembrane protein, PERK, also induced by ER stress, phosphorylates the translation initiation factor eIF2α, linking activation of the UPRER with the ISR. In the mitochondria, the UPRmt response involves ATFS1, which contains a mitochondrial targeting sequence and a nuclear localiza tion signal. Under normal cellular conditions, ATFS1 is imported into mitochondria and degraded, but upon mitochondria stress, ATFS1 is directed only to the nucleus to regulate transcription of genes encod ing mitochondrial chaperones, mitochondrial import machinery, and glycolysis. In mammals, the UPRmt is regulated by ATF5, which is the orthologue of ATFS1 in Caenorhabditis elegans. Inhibitors

of mito chondrial electron transport in mammals also activate the ISR through the release of a protease OMA1 that cleaves a cytosolic protein DELE1 to activate HRI. The ISR is induced by one or more of four kinases: PKR, PERK, HRI, or GCN2. All four of these kinases phosphorylate eIF2a, a key protein in the ternary complex that regulates protein synthesis. The resulting global inhibition of protein synthesis paradoxically pro motes translation of mRNA molecules with specific sequences in their upstream open reading frames. These include the transcription factor ATF4, which induces the expression of a gene program that maintains metabolism to preserve stress resilience. Activation of ATF4 also induces its expression and the expression of Ddit3 (CHOP), lowering the threshold for apoptosis, and Gadd45a, a phosphatase that dephos phorylates eIF2a to restore translation. The ISR is linked to ER stress through PERK and to mitochondrial stress through HRI. GCN2 is activated by amino acid deprivation, and PKR is activated during viral infection. Of particular note has been the development of ISRIB, a small molecule that partially inhibits the activity of the ISR with salu tary effects across diverse animal models of age-related degenerative diseases. These findings suggest that cell stress pathways that are pro teome protective in youth might become pathologic in aging, making them attractive targets for therapeutic intervention. Indeed, activation of the ISR has been shown to impair stem cell differentiation, perhaps linking mitochondrial dysfunction with aging, proteostasis, and stem cell dysfunction. In metazoans, the integration of stress survival strategies includes the antioxidant factor SKN-1/NRF2, the insulin-signaling factor DAF16/FOXO, and the tissue identity factor PHA-4/FOXA. Oxidative and xenobiotic stresses activate OxR, which controls the expression of redox-regulatory proteins and components of protein degrada tive pathways mediated in mammals by NRF1/NFE2L1 and NRF2/ NFE2L2, which corresponds to SKN-1 in C. elegans. NRF1 is an ERresident factor that undergoes regulated proteolytic cleavage upon activation to control expression of genes encoding subunits of the pro teasome and the UPS. NRF2 in the cytoplasm is negatively regulated by the redox-sensitive ubiquitin ligase KEAP-1; consequently, inactivation of KEAP-1 by oxidative and electrophilic stress leads to stabilization and nuclear translocation of NRF2, which in turn induces the expres sion of antioxidant proteins and detoxification enzymes. ORGANISMAL PROTEOSTASIS IN

AGING AND DISEASE Much of our understanding of protein quality control mechanisms has come from in vitro studies with purified molecular chaperones or com ponents of the UPS, complemented with cell extracts and cell-based assays using yeast or mammalian tissue culture cells. A common theme that emerges from these studies is that of hormesis, in which chronic low-level activation of the HSR, UPRER, and UPRmt is protective against subsequent exposures to extreme and lethal cell stress conditions. The importance of these pathways is further highlighted by studies in metazoans that indicate that cell stress responses are regulated at the organismal level by neuronal signaling. At the organismal level in

C. elegans, the HSR, UPRER, and UPRmt are regulated by cell-nonauton omous control by specific sensory neurons. When neuronal signaling is impaired, the HSR reverts to cell-autonomous control. Likewise, neu ronal signaling regulates the UPRmt with disruption of mitochondrial function in C. elegans neurons activating the UPRmt in nonneuronal tis sues, supporting a role for a mitokine signal. Perturbation of the mito chondrial electron transfer chain (ETC) was shown to increase lifespan in both invertebrates and rodents through the activation of the UPRmt. The response to mitochondrial dysfunction in C. elegans depends on the severity of mitochondrial impairment, with a mild reduction of ETC or reduction of mitochondrial proteostasis having hormetic effects on organismal stress resilience, proteostasis, and longevity by resetting the cytoplasmic HSR through

HSF-1, independent of ATFS-1 and the UPRmt. Mild perturbation of the ETC in Drosophila muscle also has systemic benefits on organismal health and lifespan involving the insulin signaling pathway. Communication between neurons also regulates the UPRER in peripheral tissues of C. elegans. During infection of C. elegans by pathogens, induction of the UPRER in nonneuronal tissues is mediated by sensory neurons, suggesting an organismal stress response. Cell-nonautonomous regulation of the UPRER has also been observed in mice, where overexpression of active XBP1 in pro-opiomelanocortin neurons activates the UPRER in the liver. Several other forms of intertissue communication regulate proteos tasis with beneficial effects on organismal health in model organisms. For example, muscle cell proteostasis in C. elegans is regulated by cho linergic signaling across the synaptic junction through modulation of HSF-1 activity. Transcellular chaperone signaling between somatic tis sues, and between somatic tissues and neurons, of C. elegans communi cate proteotoxic stress signals via the tissue code factor PHA-4/FOXA to control systemic expression of HSP90. In Drosophila, overexpression of small HSPs only in the flight motor muscle cells protects neurons and glial cells from elevated temperature-induced death. Enhanced expression of DAF-16/FOXO in the intestine enhances proteostasis in distant muscle cells of C. elegans, and likewise, overexpression of dFOXO/4E-BP in Drosophila muscle influences proteostasis in retina, brain, and adipose tissues to delay the age-dependent accumulation of protein aggregates. Cell stress responses and proteostasis decline in aging with insights on the relationships between processes coming primarily from studies using C. elegans with support from other invertebrate and vertebrate model systems and human cells. Endogenous metastable proteins that harbor temperature-sensitive properties misfold in C. elegans at the permissive temperature in early aging associated with a decline in the HSR. This functional decline of proteostasis in C. elegans aging is regulated by cell-nonautonomous control, from the germline stem cells to the somatic tissues for the programmed repression of the organismal HSR, resulting in the loss of stress resilience and proteostasis caus ing age-associated protein aggregation. This HSR repressive switch is regulated by signaling from the germline stem cells to the somatic tissues, resulting in the timed placement of repressive H3K27me3 chromatin marks at the promoters of heat shock genes, causing chro matin inaccessibility for HSF-1. This age-dependent decline in the HSR can be reversed either by blocking the signal from germline stem cell signal(s) or preventing the epigenetic repressive marks. The rela tionship between reproduction and inducibility of the HSR observed in animals at reproductive maturity suggests that the age-associated events of cellular failure and loss of tissue robustness during aging are not random processes but rather highly regulated, perhaps to ensure that somatic tissues are programmed to decline after reproduction, consistent with the germline soma theory of aging. Proteostasis represents one of the primary hallmarks of the biology of aging, which together with genomic instability, telomere attrition, epi genetic alterations, deregulated nutrient sensing, mitochondrial dysfunc tion, cellular senescence, stem cell exhaustion, and altered intercellular communication provides a mechanistic basis for the aging process. The programmed decline of proteostasis in early adulthood would support the hypothesis that failure in protein quality control would have negative consequences on the other pillars of geroscience. Whether proteostasis

collapse is the first to fail or among the earliest events that fail in aging, it is consistent with very large number of human degenerative diseases in aging associated with protein misfolding.

PROPERTIES OF PROTEIN

FOLDING DISEASES The complexity that arises with protein folding diseases is that all tissues are at risk and all proteins are at risk for misfolding and lossof-function or gain-of-function proteotoxicity. Added to this is the effect of aging and that each protein folding disease exhibits a highly variable age of onset for pathology. There is additional complexity in classification regarding whether to organize folding diseases by tissues (i.e., muscle proteinopathies or neurodegenerative diseases), according to the specific protein that misfolds such as α1-antitrypsin deficiency, or by the biophysical nature of the misfolded or aggregated species in amyloidoses. Disorders in which a specific mutation leads to protein misfolding or the formation of a specific insoluble protein aggregate likely repre sent only the tip of the iceberg of protein folding disorders. Mutations in aggregation-prone proteins coupled with changes in the cellular environment and effects on the capacity of the PN will promote mis folding and aggregation in affected tissues. Chronic stress may cause the aberrant cell stress responses and protein quality control pathways, causing further collateral damage and aggregation of other at-risk proteins. Such a mechanism may only manifest clinically after a seem ingly random systemic stress like pneumonia, large bone fracture, or ischemic vascular event, possibly explaining the rapid (1–2 years) accumulation of age-related morbidities in the year following a major biologic stress, marked by a need for hospitalization. As such, the age-related decline in the function of any of the components of the PN could underlie the compounding multiple morbidity that limits health span and lifespan in many elderly individuals. Within this framework, it is useful to discuss some of the better understood mechanisms of proteostasis dysfunction that have been causally linked to diseases in humans. CHAPTER 504 Protein Folding Disorders ■ ■DISORDERS THAT ENHANCE CLIENT MISFOLDING AND CAUSE PREMATURE DEGRADATION (CYSTIC FIBROSIS) Cystic fibrosis (CF) is a recessive disorder caused by mutations in both alleles of the cystic fibrosis transmembrane conductance regulator (CFTR) gene that encodes a multidomain membrane-spanning chlo ride ion channel protein. Thousands of mutations in CFTR have been identified that affect CFTR biosynthesis, folding, trafficking, and func tion, leading to chronic obstructive lung disease, intestinal obstruction, liver dysfunction, exocrine and endocrine pancreatic dysfunction, and male infertility. CF is a folding disease due to its recognition by the PN as misfolded protein. The most prominent mutation is deletion of phenylalanine 508 (F508del), present in ~90% of CF patients. Mutant ΔF508 retains partial channel function, but because it is recognized as misfolded in the ER and the cytoplasm, it is marked with ubiquitin for degradation by the UPS. Combinations of small molecules that protect the misfolded CFTR protein from degradation and enhance its func tion have led to substantial improved outcomes in many patients with CF (Chap. 302). ■ ■DISORDERS THAT INDUCE TOXIC AGGREGATES AND LOSS OF FUNCTION IN MULTIPLE TISSUES (`1-ANTITRYPSIN DEFICIENCY) α1-Antitrypsin deficiency (AATD) is a co-dominant inherited disease with an increased risk of chronic obstructive pulmonary disease, liver disease, and inflammation of the blood vessels. Pulmonary problems are more frequent in adults, whereas liver and skin problems may occur in adults and children. α1-Antitrypsin is encoded by the SERPINA1 gene and secreted into the circulation by the liver and is responsible for inactivating endogenous proteases, particularly those secreted by neutrophils and other inflammatory cells in the lung. Patients with AATD harbor mutations in SERPINA1 that cause misfolding in the ER. The two major phenotypes resulting from this abnormality highlight

the diverse consequences of misfolding on different cells and organs. In the liver, misfolding of the mutant protein results in the formation of toxic aggregates and hepatocyte death, manifest as liver injury and eventually cirrhosis—a gain-of-function toxicity. In the lung, the fail ure to secrete

sufficient α1 antitrypsin may lead to unchecked proteo lytic damage to the delicate architecture of the alveolus, a process that is markedly worsened when neutrophils are recruited to the lung in response to cigarette smoking. This loss-of-function phenotype mani fests pathologically as emphysema and clinically as chronic obstructive pulmonary disease.

■ ■INTERACTIONS WITH PN COMPONENTS

THAT CHANGE CONFORMATION, STABILITY,

OR FUNCTION (CANCER) Mutations in the tumor suppressor p53 are among the most com mon mutations observed in patients with cancer. Deletion of p53 combined with overexpression of an oncogene is sufficient to drive metastatic cancer formation in mice, causally linking p53 mutations with cancer. Normally, p53 functions as a transcription factor that suppresses the transcription of genes involved in apoptosis resistance. While myriad mutations in p53 have been described, some result in an alternate conformation that interacts with different HSP70 chap erones within the PN. Binding of the mutant p53 protein to these chaperones affects the DNA binding property necessary for its tumor suppressor function and facilitates binding to other domains, result ing in changes in gene expression that protect malignant cells from apoptosis. ■ ■STRONGLY ENHANCED AGGREGATION PROPENSITY AND AMYLOID FORMATION (ALZHEIMER’S DISEASE, PARKINSON’S DISEASE, AMYOTROPHIC LATERAL SCLEROSIS, HUNTINGTON’S DISEASE, TYPE 2 DIABETES MELLITUS) In some individuals, native or mutant proteins include sequence motifs that promote a highly ordered aggregation state when the cellular environment is altered. The most common of these motifs is the beta sheet, which, when exposed to the solvent environment of the cell, forms intermolecular species that bind in an iterative process that can accommodate many thousands of molecules that form cross-beta sheet amyloids. These intracellular aggregates are described as oligomers (2–24 molecules), protofibrils (rods 4–11 nm wide and 200 nm long), and amyloid fibrils with a similar width to protofibrils but microns in length. While the formation of oligo mers is thermodynamically unfavorable, polymerization is favorable, causing aggregates to seed slowly but grow exponentially. In some cases, for example in Huntington’s disease and familial forms of Alzheimer’s disease and ALS, aggregation is accelerated by mutations or expansion of homopolymers. However, in many cases, the aggre gates contain other cellular proteins that share biophysical proper ties of aggregation propensity or reflect dysfunction in the PN that facilitates their seeding or propagation (see below). While in most instances, damage caused by protein aggregates is localized to the cells in which they form, as occurs with islet amyloid peptide in some patients with type 2 diabetes mellitus, amyloidogenic proteins asso ciated with neurodegenerative diseases have been shown to spread between cells and, in the case of transthyretin amyloidosis, can cause pathology in many tissues. Pathologists use staining of tissues with Congo red, which detects beta sheets, to make this diagnosis. Dam age to neurons by aggregates in Alzheimer’s disease can elicit a local inflammatory response by resident immune cells in the brain, both of which contribute to pathology. Much effort has been directed toward the detection of aggregates and amyloid and the development of small molecules or antibodies that block further growth or enhancing the cellular activities of the PN to suppress protein misfolding. PART 20 Emerging Topics in Clinical Medicine ■ ■SECRETED AGGREGATED AND AMYLOID SPECIES CAUSING SYSTEMIC AMYLOIDOSIS In patients with systemic amyloidosis, the secretion of large amounts of aggregation-prone proteins results in the deposition of aggregates in

many tissues. These proteins can include immunoglobulins secreted from plasma cells in patients with systemic inflammation or multiple myeloma or other aggregation-prone proteins including transthyretin. Similar to other aggregate-induced diseases, mutations in transthyretin that enhance polymerization are associated with an increased risk of developing systemic amyloidosis with advancing age. These aggregates induce cellular toxicity, inflammation, and matrix reorganization, which interfere with function in an organ-specific manner. ■ ■NATIVE PROTEINS PRONE TO AGGREGATE

WHEN THE CELLULAR ENVIRONMENT IS

ALTERED BY STRESS AND AGING While well-defined genetic abnormalities have been essential in elucidating the molecular mechanisms that underlie the formation of protein aggregates and causally linking them to disease, many if not most clinical diseases associated with the formation of protein aggregates develop in patients without identified mutations. In these patients, a decline in the chaperone and quality control mechanisms of the PN allows exposure of aggregation-prone domains of normal proteins to the solvent environment of the cell. Once seeded, these protein aggregates can expand rapidly to induce local or systemic injury. The decline in function of the PN that allows these aggregates to form might develop gradually with advancing age or might occur suddenly in response to an age-triggered biologic program, as occurs in C. elegans. ■ ■INFECTIOUS DISEASES AND IMBALANCED

CELL STRESS RESPONSES IN AGING The response to infectious diseases and the disproportionate morbidity and mortality in older individuals exposed to systemic stress are likely associated with the decline in robustness of cell stress responses and proteostasis. While these stressors include infections, surgical or acci dental trauma, sepsis, and myocardial infarction, among others, pneu monia, the most common cause of death from an infectious disease in the United States, provides an illustrative example. As was evident during the COVID-19 pandemic, pneumonia morbidity and mortal ity disproportionately affect the elderly. Viral pneumonias, including those caused by influenza viruses and SAR-CoV-2, are primarily local ized to the lung, where they activate a local and systemic inflammatory response and denude the alveolar lining. The resulting hypoxemia and systemic inflammatory response injure distant organs independent of viral injury. Impaired function of the PN during the stress might allow seeding of tissues with toxic aggregates with long-term conse quences. Repair of the damaged lung and distant organs represents a major challenge to proteostasis that might be overcome in younger individuals but fails in those who are older with poor stress resilience. This loss of proteostasis resilience necessary to limit damage and allow repair could explain clinical observations in pneumonia survivors who develop persistent lung injury, skeletal muscle dysfunction impairing mobility, chronic kidney disease, cognitive dysfunction, and dementia and an increased risk of ischemic cardiovascular events in the year after hospital discharge. ■ ■FURTHER READING Balch WE et al: Adapting proteostasis for disease intervention. Science 319:916, 2008. Balchin D et al: In vivo aspects of protein folding and quality control. Science 353:aac4354, 2016. Chiti F, Dobson CM: Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. Annu Rev Biochem 86:27, 2017. Costa-Mattioli M, Walter P: The integrated stress response: From mechanism to disease. Science 368:eaat5314, 2020. Eisele YS et al: Targeting protein aggregation for the treatment of degenerative diseases. Nat Rev Drug Discov 14:759, 2015. Finley D, Prado MA: The proteasome and its network: Engineering for adaptability. Cold Spring Harb Perspect Biol 12:a033985, 2020.

Labbadia J, Morimoto RI: The biology of proteostasis in aging and disease. Annu Rev Biochem 84:43, 2015.

13 - 505 Novel Approaches to Diseases of Unknown Etiology

505 Novel Approaches to Diseases of Unknown Etiology

Levine B, Kroemer G: Biological functions of autophagy genes: A disease perspective. Cell 176:11, 2019. Mallucci GR et al: Developing therapies for neurodegenerative disor ders: Insights from protein aggregation and cellular stress responses. Annu Rev Cell Dev Biol 36:165, 2020. Song J et al: Quality control of the mitochondrial proteome. Nat Rev Mol Cell Biol 22:54, 2021. David R. Adams, Camilo Toro, Joseph Loscalzo

Novel Approaches to

Diseases of Unknown

Etiology THE UNDIAGNOSED DISEASE STATE The term disease, etymologically meaning “lack of ease” or the presence of discomfort, is defined as an abnormal state that negatively affects the structure or function of all or part of an organism and that is not due to any immediate external injury. When referring to a person experienc ing a disease, the word patient is used in its original, meaning “the one who endures suffering.” These terms are well suited to patients with undiagnosed diseases. A patient with an undiagnosed disease is one for whom a medical diagnosis is not discerned after reasonable efforts utilizing established methods and procedures. Multiple factors may contribute to a failure to reach a diagnosis (Table 505-1). Patients who are affected by an undiagnosed disease for a protracted period exist in an undiagnosed state, which presents characteristic challenges for the patients, their families, and their medical providers. ■ ■THE MEANING AND CONTEXT OF A DIAGNOSIS A diagnosis often entails hierarchical levels of information specific ity with varying levels of relevance to the users (consumers) of such information (e.g.) patients and their families, health care providers, government health care agencies, insurers, epidemiologists, genetic counselors, pharmacologists, biologists). As an example, a diagnosis of Parkinson’s disease in an adult is based on the progressive emer gence of signs and symptoms of bradykinesia, rigidity, asymmetric rest tremor, and postural instability (clinical diagnosis), which are typically responsive to the administration of L-dopa (a therapeutic

response biomarker). Together, these are cardinal features of striatonigral degen eration (a mechanistic diagnosis), a process associated with neuronal α-synuclein deposition and Lewy body pathology (histopathologic diagnosis) often based on a genetic susceptibility conferred by muta tions in genes such as synuclein (SYNCA, a molecular diagnosis) and likely influenced by environmental exposures (e.g., manganese or other neurotoxins). With ongoing advances in medical science and technology, the standard for what constitutes a reasonable diagnosis continues to evolve toward higher levels of specificity. For example, the utility of an antisense oligonucleotide therapy may be restricted to a specific subset of mutations associated with a given, monogenic, heritable disease. Efforts to adopt the principles of precision medicine include a growing emphasis on the context of disease within the genomic landscape, envi ronment, social factors, medical history, nutrition, and the microbiome of any given individual. Examples include cancer susceptibility, geneti cally determined idiosyncratic reactions to medications, and unique pathogen susceptibilities in patients with certain immune deficiencies. ■ ■UNDIAGNOSED RARE DISEASES Most chronically undiagnosed diseases are rare. While individual rare diseases have a low prevalence, they are numerous in aggregate. It is

TABLE 505-1 Factors Contributing to the Presence of an

Undiagnosed Disease FACTOR EXAMPLE Misleading information False-negative and false-positive test results Rare disorder Many inherited disorders have only been identified in a few individuals. For example, sialuria, a well-understood disorder of sialic acid metabolism, has been reported in 10 individuals (OMIM 269921). Unusual causes of common diseases, including atypical course of illness Insulin-dependent diabetes mellitus may be the presenting feature for the relatively rare autoimmune polyendocrinopathy syndrome, type I (OMIM 240300). Presence of multiple disorders (blended phenotypes) For an example, see PubMed ID 24863970. Lack of characteristic symptoms of known disease Diseases are commonly ascertained via cardinal signs or symptoms leading to incomplete ascertainment of all possible disease presentations. For instance, not all persons with Marfan’s syndrome are tall relative to other family members. For progressive diseases, pathognomic signs and symptoms may be missing in early stages of disease. New disease No prior knowledge or record of such disease Incorrect affected status assignments in family history A heritable disorder may be inappropriately excluded if family history information is incorrect. Primary disease manifestations obscured by other factors Maladaptive behavior, medication effects, and secondary disease manifestations may obscure signs and symptoms of a primary disorder. CHAPTER 505 Disease not expected in region or population Cystic fibrosis in persons of African ancestry, sickle cell disease in persons of northern European ancestry; infectious agents with marked geographical incidence patterns Diseases thought to be eradicated Poliomyelitis Novel Approaches to Diseases of Unknown Etiology
Diseases occurring in unexpected time of life Parkinson’s disease in children, lysosomal storage disease in adults Malingering Feigned disease features intended to achieve secondary gain (Munchausen syndrome) Rare disease mechanisms Transmitted or sporadic prion disease, unusual zoonotic diseases Abbreviation: OMIM, Online Mendelian Inheritance in Man. estimated that >6000 rare diseases affect millions of people throughout the world. Estimates of aggregate population prevalence range from 6 to 10%. Many rare diseases have a genetic basis and onset in child hood. As the cloud of uncertainty inherent in the undiagnosed disease state is removed, new disease-specific counseling, therapies, resources, community engagement, and advocacy opportunities become possible. ■ ■THE EFFECT OF THE UNDIAGNOSED

DISEASE STATE ON THE PATIENT Patients with an undiagnosed disease are frequently driven to under stand the basic nature of their ailment (what, when, where, how, etc.). Individuals, families, physicians, and society, however, might have a wide range of tolerance to the uncertainties associated with the undi agnosed disease state. Being undiagnosed has profound detrimental effects. Patients can go undiagnosed for decades, leading to personal and family uncertainty, high levels of stress, decreased productiv ity, limited accessibility to disease-specific counseling and resources, decreased quality of life, and excess utilization of medical services. APPROACH TO CHALLENGING DISEASES OF UNKNOWN ETIOLOGY Approaches to a patient with an undiagnosed disease can be separated into two categories. The first is a new assessment by a consultant, new provider, or diagnostic referral center. The second is periodic reassess ment by an existing provider for a patient who remains undiagnosed.

TABLE 505-2 Essential Records for Undiagnosed Disease Patients

1. Any narrative summaries that detail the course of the illness
2. Copies of original test results with names, dates, testing circumstances, normal ranges, and test facility information
3. Electronic copies of imaging studies
4. Consultation notes
5. Hospitalization intake and discharge summaries
6. Accurate family history accounts and family relations Optional but potentially useful records include:
7. Photographs and/or videos of disease manifestations
8. Longitudinal data (growth charts, symptom logs, serial lab measurements)
9. Data or specimens that could be reanalyzed, including pathology specimens and genomic sequencing of raw data ■ ■COMPREHENSIVE DATA COLLECTION A potentially time-consuming but critical initial step is the gathering of all available medical data. Essential records are listed in Table 505-2. The overall goal of data collection is a full understanding of the course of the disease and a verification of critical data elements used for diagnostic decision-making. Incorrect or partial second-hand accounts of prior test results contribute substantively to incorrect or missed diagnoses. Analysis of the collected data allows for reconstruction of the process by which previous disease presentation, diagnostic thought processes, and test interpretation led to the current understanding of a patient’s illness. Unintentional obfuscation of the history and findings can result from missing records, incomplete recall by the patient and fragmentation, and propagation of information (and misinformation) in the medical record. Optimally, the presence and character of key features of the illness will be reinforced by perspectives derived from multiple evaluations. PART 20 Emerging Topics in Clinical Medicine ■ ■VALIDATION OF SUBJECTIVE AND

OBJECTIVE FINDINGS Teasing apart the layers of a patient’s presentation often uncovers a variety of adaptive (and maladaptive) coping strategies. Some are idiosyncratic to the disease state (e.g., sun avoidance in a patient with xeroderma pigmentosum), whereas others are driven by psychoso cial factors and could become primary drivers of the phenotype. It is important to consider, however, that patients believed to have “functional” or “somatoform” disorders may have unrecognized underlying ill nesses, e.g., nonepileptic events (pseudo-seizures), frequently have concurrent bona fide epileptic events. Careful consideration of clinical phenomenology and

associated findings on physical examination and ancillary investigations may provide clarity, affirmation, and effective redirection. Distinct clinical, radiographic, and laboratory abnormali ties provide entry points to the generation of a differential diagnosis and could become effective biomarkers of disease progression and response to interventions. Testing Strategies and New Technology The historical exclu sion of a diagnostic hypothesis may be based on testing that is no lon ger state of the art. For example, congenital disorders of glycosylation (CDG) were historically diagnosed using transferrin isoelectric focus ing. It was subsequently found that the diagnosis of many CDG types required mass spectrometric and molecular approaches. The initial assessment of a patient with an undiagnosed disease should include a reassessment of the diagnostic logic and data used in past decisionmaking. In the absence of concrete diagnostic leads, the use of broad scope screening tools may prove beneficial in generating meaningful diagnostic hypotheses (Table 505-3). In some cases, newer testing options may be difficult to obtain and/or be costly. Prior probability of disease and available resources will factor in determining whether new diagnostic testing is practical. Molecular Approaches, Including Genomics The availability and variety of clinical molecular modalities have transformed diag nostic testing in many settings. These advances have arisen from both

TABLE 505-3 Clinically Available Tests with Notable Utility for Undiagnosed Cases TEST TARGET PHENOTYPES RATIONALE Single nucleotide polymorphism microarray and/or karyotype Dysmorphic features, cognitive impairment, neurodevelopmental disorders Genomic structure abnormalities may be missed by other testing Exome or genome sequencing Any undiagnosed disease that is chronic and not clearly acquired Tests a broad range of potentially unconsidered diagnostic entities Lysosomal storage diseases (LSDs), molecular or enzymatic tests. urine organic acids, urinary glycosaminoglycans (GAGs), oxysterols Progressive neurologic disorders, psychiatric disorders Some LSDs have nonspecific presentations, and adultonset cases are often missed Congenital disorders of glycosylation, Apo CIII and N-glycan mass spectrometry Pediatric-onset disorders, cognitive impairment, neurologic phenotypes Large group of disorders; phenotypes for many still being characterized Biochemical disorders, ammonia, serum polyols, urine purines and pyrimidines, plasma amino acids, very-longchain fatty acids Neurologic phenotypes, especially with waxing and waning course, selective speech involvement, or patients with unusual selfselected diets Metabolic disorders may have nonspecific symptoms, and adultonset cases are often missed Mitochondrial sequencing and mitochondrial depletion studies; biochemical screening with serum lactate, blood pyruvate, plasma amino acids, and GDF-15 Complex multisystem disorders with neurometabolic, endocrine, and gastrointestinal symptoms, muscle dysfunction, and waxing and waning or progressive course Large group of disorders with a wide range of presentations; yield is improved by studies in affected tissue (e.g., liver or muscle) Cerebrospinal fluid (CSF) studies including amino acids (AAs), lactate, pterins, methyltetrahydrofolate (MTHF), or special CSF flow studies Synthetic neurotransmitter defects in patients with unexplained fluctuating encephalopathy/ movement disorders or patients with atypical neuroinflammatory syndromes Patterns of profiles point to particular enzymatic deficits in neurotransmitter synthesis or characterization of unique immunologic profiles of inflammatory central nervous system diseases testing scope, such as exome-wide, genome-wide, and transcriptome (RNA sequencing [RNA-Seq]) sequencing, and new medical knowl edge, such as new disease-gene associations and molecular interaction networking (network medicine). Complementary screening tools, such as metabolomics, show diagnostic promise, particularly when combined with sequencing data to generate a fuller picture of disease manifestations. Simultaneous consideration of multiple data types can provide a means of

appreciating overlapping and reinforcing evidence, with the potential to inform both hypothesis-driven and agnostic approaches to diagnosis. HYPOTHESIS-DRIVEN MOLECULAR TESTING Hypothesis-driven test ing implies that a defined set of heritable (or potentially heritable) dis orders is the principal impetus for testing. Selection of a targeted gene sequencing panel, ideally augmented with structural variant detection, may allow for improved sensitivity, lower cost, and fewer unrelated (secondary) findings relative to full exome or genome sequencing stud ies. In the setting of an initial undiagnosed disease evaluation, prior sequencing panels may not include recently discovered genes. Testing with an updated panel or targeted sequencing of a newer gene is an option for consideration. In some cases, sequencing panels are gener ated by selective reporting of relevant genes within an exome dataset. In such cases, it may be possible to expand the analyses to new genes of interest without additional sequencing.

AGNOSTIC MOLECULAR TESTING Agnostic testing typically uses data from a broad testing platform, such as exome or genome sequencing, and considers all detectable diagnoses, even those with a low pretest probability of being present. This approach can also generate hypoth eses for potentially new disease-gene associations. Analysis of the sequencing data typically includes an unrestricted search throughout the entire human genome or exome space. DNA sequence variants with potential medical relevance are identified first by bioinformatic characteristics, including known association with disease, predicted importance for protein function, interspecies conservation, popula tion frequency, and an evolving list of other associated or functional factors. The list of candidate variants is then subject to expert review (i.e., curation). The interpretation of test results in this setting is highly influenced by both the adequacy of communication between the clini cal and testing teams and the information content of the data sources used to annotate each of the thousands of variants generated in the course of sequencing. There is a rapid proliferation of new testing platforms and analytical tools with the potential to contribute to solving undiagnosed diseases, but it remains challenging to judge their broad utility. While awaiting systematic validation and practice standards, novel techniques may be considered in special cases where a diagnostic hypothesis is closely aligned with the type of data generated by a specific testing strategy (Table 505-4). ■ ■PERIODIC REEVALUATION The cornerstone for the care of a patient in an undiagnosed disease state is a plan for periodic reevaluation. The Undiagnosed Diseases Network (UDN), a 10-year National Institutes of Health–sponsored national program, was specifically designed to evaluate undiagnosed patients. The overall diagnostic rate of the UDN, including periodic reevaluation of early enrollees, was reported as ~30%. This finding TABLE 505-4 Emerging or Special Testing Strategies and Related Diagnostic Questions AVAILABLE CLINICALLYa TESTING STRATEGY RELATED DIAGNOSTIC QUESTION Transcriptomics, RNA-Seq Relevance of splice, regulatory, and other noncoding variants; correlated changes in gene expression within pathways Yes Metabolomics Hypothesis generation via nontargeted approaches, correlated pathway changes, correlation with molecular findings Yes Epigenetics Diseases known or suspected to be caused by methylation or parent-of-origin effects Yes Transcriptional profiling Search for profile particular to certain disease states, e.g., interferon-inducible gene panels (interferon signature) in certain autoinflammatory disorders Some Specialized, diseasespecific testing Prion-related diseases, metabolic diseases, and many other assays Some Functional validation Model organisms, cell biology, and other approaches to validating a hypothesized gene-disease association No Metagenomics Search for molecular fingerprints of other organisms (e.g., infectious agents) within human samples Yes Long-read sequencing technology Accurate resolution of low-complexity regions of the human genome (repeat expansion disorders) and complex genome

structural rearrangements No Deep sequencing Accurate resolution of low levels of mosaicism Some Optical genome mapping High-resolution chromosomal structure Yes aAvailable clinical tests are often a small subset of approaches available via research collaboration. Clinical testing offerings are evolving rapidly and should be reassessed periodically.

illustrates the fact that many affected individuals remain in an undi agnosed state for a protracted period. For a medical provider, the care of an undiagnosed patient includes a program for symptomatic care, support related to the undiagnosed state itself, and plans for a regular reevaluation strategy seeking new insights into the diagnosis by follow ing its time trajectory. Reevaluation is guided by emerging knowledge in the field, disease progression, and the development of new signs and symptoms. The appearance of a similar disease in a sibling or close relative may provide critical insight. Communication with the patient is an essential component. Many individuals with an undiagnosed dis ease report feeling abandoned by their providers once initial diagnostic ideas have been exhausted. Providers themselves may feel discouraged in being unable to provide a diagnosis. The institution and discussion of a well-defined plan for periodic reassessment and communication can help reinforce the patient-provider relationship and set reasonable expectations.

Reevaluation of Differential Diagnosis The key to success for a planned reevaluation visit is preparation. The problem and differential diagnosis lists should be subject to careful, evidence-based review. New or resolved clinical features may add or remove diagnostic consider ations. The passage of time may result in the emergence of distinct new phenotypic manifestations that serve as new clues in the formulation of a definitive diagnosis. Special consideration should be given to the effects of reaching maturity and aging. The establishment of a pheno type as being static versus progressive has prognostic value. Careful documentation of the rationale for including or excluding individual disorders will streamline the process for both future reevaluations and the need for consultants. Concurrent development of common diseases should be thoughtfully considered as a possible component of the primary undiagnosed condition. For example, emergence of insulindependent diabetes mellitus in an undiagnosed patient with a complex phenotype could be a feature of the rare autoimmune polyendocrinopa thy associated with mutations in the autoimmune regulator gene AIRE. CHAPTER 505 Novel Approaches to Diseases of Unknown Etiology
New Literature Keeping abreast of current literature is an impor tant and challenging activity for all medical providers as the body of medical knowledge continues to grow exponentially. For undiagnosed diseases, newly reported disorders and disease-gene associations are an important source of diagnostic resolution. Literature search tools such as PubMed can be augmented by online resources that connect clinical signs and symptoms (phenotypes) to disorders. For example, using the search terms “cardiomyopathy arthropathy diabetes hyperpigmenta tion” in the Online Mendelian Inheritance in Man website (https:// omim.org) produces a list of disorders that includes hemochromatosis. In the context of an undiagnosed disease, this type of phenotypedriven approach can be used to search for new, relevant publications and disorders. Tools for search automation continue to be developed in both open-source and commercial settings. The success of these approaches is augmented by iterative application, ideally as part of formal, periodic reevaluation of the undiagnosed patient. ■ ■GENOMICS The use of medical testing based on the determination of DNA sequence and structure (sometimes referred to as molecular testing) has proliferated in recent years. A wide variety of approaches are available to the clinician, from single-

gene sequencing to exome or genome sequencing. RNA sequencing and optical genome mapping have recently become available as clinical tests. Many reviews of this topic are available (see Marwaha et al., 2022, in “Further Reading”). Consultation with colleagues trained in genetics can be useful when developing an optimal testing approach. In some cases, genetic testing results may already exist in the medi cal record during the initial evaluation of an undiagnosed patient. This is increasingly true for younger patients; exome and genome sequenc ing are being used earlier, and with increasing frequency, for complex diagnostic challenges. Reanalysis of previously obtained exome and genome data should start with consideration of both the age and qual ity of the study and the reported patient phenotype at the time of the study report. For sequence results generated in a clinical laboratory,

a discussion between the provider and the laboratory director often answers important questions about recommended next steps. The dis cussion should touch on how technologic advances have affected the utility of the older data and whether the laboratory offers reanalysis of the data. At a minimum, the provider, the testing laboratory, or an identified subspecialist should review previously reported DNA vari ants of unknown significance considering interval reports about the gene in question. More advanced reanalysis strategies are emerging and may be offered by the testing laboratory.

Some laboratories offer release of raw DNA sequencing data to their patients on request. The utility of raw data varies and depends on the identification of bioinformatics collaborators willing to reanalyze the data. Sequencing data obtained as part of a research study may not be suitable for clinical diagnostic purposes. In practice, raw and research-generated sequencing data are most useful when a collaborating researcher can be identified. When considering a new sequencing test, the inclusion of the par ents and siblings of the proband has the potential to provide enormous value in some situations. Discussion of an optimal approach with an expert colleague or the testing laboratory is encouraged. ■ ■EXPOSOME In many cases, a detailed occupational and environmental expo sure history should be obtained. Some rare disease phenotypes are pathognomonic of specific toxicant exposures (e.g., mesothelioma and asbestos exposure, clear cell adenocarcinoma of the vagina and intrauterine diethylstilbestrol [DES] exposure, chloracne and exposure to halogenated aromatic hydrocarbons). For the most part, however, chemical toxicant exposures do not produce unique phenotypes. Rather, chemical exposures operate in conjunction with lifestyle factors (e.g., smoking, alcohol intake, and nutritional status), dif ferential host susceptibility (determined by age, sex, comorbidities, genetics, etc.), and nonchemical stressors (e.g., psychosocial stress) to produce (1) common, readily diagnosed medical diseases (e.g., asthma); (2) unusual or nonspecific phenotypes (e.g., erethism and metallic mercury exposure); or (3) atypical presentations of otherwise well-characterized disease states, initially considered an undiagnosed disease (e.g., manganese-induced parkinsonism). The nonspecificity typical of chemical-induced disease risk is further complicated by lack of exposure biomarkers for many common environmental toxicants (e.g., volatile organics), the short half-life of some contaminants (e.g., arsenic), and the possibility of decades long latency between exposure and disease onset (e.g., chemical carcinogenesis or dietary exposures to specific biochemical risk factors for atherothrombosis). In addition, we live in an era in which new chemicals are introduced into consumer products and the environment at a pace well beyond our capacity to characterize their toxicity. Within this context, one of the most power ful tools for ascertaining chemical-related disease risk is a systematic exposure history. Although there are no standardized instruments for

this purpose, there are published guidelines to implement exposure assessments (Goldman and Peters, 1981; see “Further Reading” below). These include a multistep approach to exposure assessment including a job history; a review of exposures at work and at home or via hobbies and recreation; ascertainment of any temporal relationship of symp toms or disease onset to work, home, or recreational activities; and the food frequency questionnaire. If this screening identifies a potential exposure or exposures of concern with respect to patient symptoms and phenotype, a second step of evaluation involves a more detailed history to identify specific suspect agents, options for quantitative environmental exposure assessment (e.g., household tap water sam pling, review of workplace Material Safety Data Sheets [MSDSs]) and biomonitoring, and etiologic plausibility for at least some aspects of the patient’s phenotype. PART 20 Emerging Topics in Clinical Medicine The traditional approach to focused external exposure assessment proposed above does not, however, provide an integrated, quantitative measure of all exposures over the life course, an exposure characteriza tion of particular interest for the risk of chronic diseases such as cancer or atherothrombosis. The exposome has been proposed as a promising means for capturing the totality of human exposure over a lifetime

(analogous to the totality of genetic exposure assessed via genomic analyses), including not only external chemical or dietary/foodome (Barabasi et al., 2020; see “Further Reading” below) exposures but also internal (e.g., metabolic, hormonal, microbiome) influences and psychosocial factors. However, techniques for measuring the expo some are in relatively early stages of development, are limited by the substantial variability in human exposure experience, and have not yet been designed to capture complex combinations commonly encoun tered in environmental or occupational exposure settings (Peters et al., 2012; Wild, 2012; Brunekreef 2013; Barabasi et al., 2020; see “Further Reading” below). This important element of assessing patients with undiagnosed disease is, however, evolving rapidly and offers the prom ise of becoming a more formal part of the evaluation of many patients with undiagnosed disease. ■ ■ENGAGEMENT OF RESEARCH APPROACHES Establishing a research collaboration for a patient with undiagnosed disease can be both challenging and rewarding. Time and effort resources are likely to limit this approach to a subset of patients with particularly compelling clinical presentations and a strong hypothesis about disease causation. The process must include early and detailed communication with the patient. Several approaches may be consid ered. Undiagnosed disease medicine, as a focus of specialized study, investment, and infrastructure, has proliferated around the world in the last decade, with numerous centers providing support for evaluat ing qualifying individuals and families. Leveraging Phenotypic and Genotypic Similarities For a patient with a rare or undiagnosed disease and distinct presenting fea tures, finding similarly affected individuals adds substantial benefits. It can encourage research, provide a community for affected patients, and improve the chances of finding commonalities in pathogenesis and therapeutic strategies. Phenotypic aggregation may also allow the patient to connect with consortia invested in related medical presenta tions. Examples include organizations dedicated to the study of related diseases such as leukodystrophies, autoinflammatory disorders, and even undiagnosed diseases; NORD, the National Organization for Rare Disorders (rarediseases.org), can be a useful starting point. The Office of Rare Disease Research within the National Center for Advancing Translational Science supports consortia under the Rare Disease Clini cal Research Network program. Building patient cohorts may also be based on specific biological mechanisms or pathways, for example, the United Mitochondrial Disease Foundation. Data Sharing The proliferation of DNA sequencing technology and the subsequent generation of many DNA variants of unknown clinical significance have

prompted the creation of data-sharing resources spe cifically designed to match similar cases submitted by clinicians and researchers around the world. For example, a clinical exome report may identify variants in a gene with a potential but unproven relation ship to the patient's presenting illness. The clinician could enter the gene name into a gene-matching database, and if the same gene name had been already entered by a different submitter, the database would flag a match and send contact information to both submitters. The matching procedure has the potential to identify additional cases of an ultrarare or newly described condition, while avoiding the shar ing of the patient's personal health information. Embellishments of this approach involve inclusion of phenotypic features, data entry by families, and specific details of sequence variants. Example systems include GeneMatcher (the most populated database for single-gene submissions) and DECIPHER (which adds expended utility for larger structural variants). As an illustration of their utilization at the time of this chapter's publication, GeneMatcher has single-gene submissions for almost 70% of all known protein coding genes. Collaboration Collaborations around undiagnosed disease patients may take many forms. Studies focusing on related medical conditions can sometimes be identified using the https://clinicaltrials. gov website, which lists many U.S. and non-U.S. clinical studies. Data bases of clinical information (e.g., this textbook, GeneReviews) can be used to identify subject matter experts for related conditions. Such

No diagnosis after comprehensive evaluation Symptomatic care, consider empiric treatments Consultative assessment Iterative assessment Comprehensive review and re-evaluation of evidence for prior conclusions Review new records, signs, symptoms, and environmental history Testing to validate key results and evaluate new diagnostic hypotheses Consider new literature and availability of new or updated testing strategies Consider hypothesisgenerating tests including genomic sequencing Reformulate differential diagnosis DIAGNOSIS? Yes No Document reasoning and supporting evidence Consider collaboration to explore hypotheses Work with patient to define concrete follow-up plan FIGURE 505-1 Approach to the patient with an undiagnosed disease.
experts can be queried about ongoing studies. In some cases, a willing ness to work with consenting families to provide biological specimens can open additional avenues for collaboration.
■ ■CHALLENGES Data Portability Obtaining specimens, data, and records for a chronically undiagnosed patient can be time-consuming and

challenging. Families may be charged fees for obtaining copies of old studies. Although continuing advances in record access are occurring, families should be encouraged to collect and maintain an updated col lection of medical records. These should include copies of consultation notes, original laboratory results, and radiology studies (the latter pref erably in electronic form). These record collections are useful for con sultation, second opinions, and transitions between primary providers.

Managing Illness Behaviors, Expectations, and Secondary Manifestations Patients with undiagnosed diseases may present in any stage of the grieving process. Coping with uncertainty, loss of abilities, work, relationships, autonomy, and financial security com pound the primary manifestation of the disease. Patients may have a wide range of expectations about the possible benefits of achieving a diagnosis, including successful therapy. Patients of reproductive age may find that their greatest uncertainty surrounds the potential herita bility of their disorder, its effects on future reproductive decisions, and the potential risk it may represent to their children and living relatives. These factors may be equally or more disabling than the primary illness and require an individualized and multidisciplinary approach. CONCLUSION Chronically undiagnosed diseases

present a complex challenge to patients, medical providers, and society at large. Development of a comprehensive plan for evaluation, reevaluation, and support requires a substantial investment of time and effort (Fig. 505-1). Achieving an accurate diagnosis removes at least one level of uncer tainty and allows for disease-specific counseling, therapies, resources, community engagement, and advocacy opportunities otherwise not afforded to undiagnosed patients. CHAPTER 505 ■ ■FURTHER READING Barabasi AL et al: The unmapped chemical complexity of our diet. Nat Food 1:33, 2020. Brunekreef B: Commentary: Exposure science, the exposome, and Novel Approaches to Diseases of Unknown Etiology

public health. Environ Mol Mutagen 54:596, 2013. Goldman RH, Peters JM: The occupational and environmental health history. JAMA 246:2831, 1981. Lee CE et al: Rare genetic diseases: Nature’s experiments on human development. iScience 23:101123, 2020. Marwaha S et al: A guide for the diagnosis of rare and un-diagnosed disease: Beyond the exome. Genome Med 14:23, 2022. Peters A et al: Understanding the link between environmental expo sures and health: Does the exposome promise too much? J Epidemiol Community Health 66:103, 2012. Splinter K et al: Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med 379:2131, 2018. Wild CP: The exposome: From concept to utility. Review. Int J Epide miol 41:24, 2012.

This page intentionally left blank