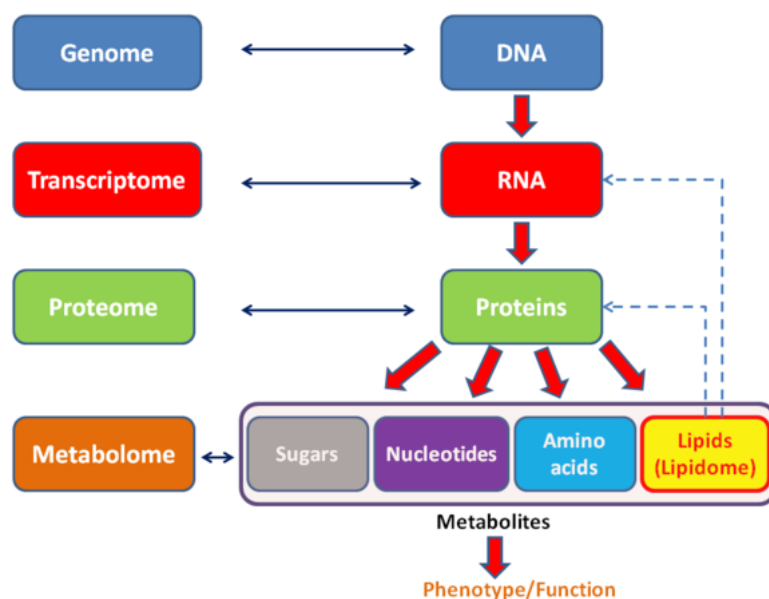


Metabolome

The **metabolome** refers to the complete set of small-molecule chemicals found within a biological sample.^[1] The biological sample can be a cell, a cellular organelle, an organ, a tissue, a tissue extract, a biofluid or an entire organism. The small molecule chemicals found in a given metabolome may include both endogenous metabolites that are naturally produced by an organism (such as amino acids, organic acids, nucleic acids, fatty acids, amines, sugars, vitamins, co-factors, pigments, antibiotics, etc.) as well as exogenous chemicals (such as drugs, environmental contaminants, food additives, toxins and other xenobiotics) that are not naturally produced by an organism.^{[2][3]}



General schema showing the relationships of the genome, transcriptome, proteome, and metabolome (lipidome).

In other words, there is both an endogenous metabolome and an exogenous metabolome. The endogenous metabolome can be further subdivided to include a "primary" and a "secondary" metabolome (particularly when referring to plant or microbial metabolomes). A primary metabolite is directly involved in the normal growth, development, and reproduction. A secondary metabolite is not directly involved in those processes, but usually has important ecological function. Secondary metabolites may include pigments, antibiotics or waste products derived from partially metabolized xenobiotics. The study of the metabolome is called metabolomics.

Contents

Origins

Measuring the metabolome

Metabolome databases

The Human Metabolome Database

Human biofluid metabolomes

Yeast Metabolome Database

The Escherichia coli Metabolome Database

See also

References

External links

Origins

The word metabolome appears to be a blending of the words "metabolite" and "chromosome". It was constructed to imply that metabolites are indirectly encoded by genes or act on genes and gene products. The term "metabolome" was first used in 1998 ^{[1][4]} and was likely coined to match with existing biological terms referring to the complete set of genes (the genome), the complete set of proteins (the proteome) and the complete set of transcripts (the transcriptome). The first book on metabolomics was published in 2003.^[5] The first journal dedicated to metabolomics (titled simply "Metabolomics") was launched in 2005 and is currently edited by Prof. Roy Goodacre. Some of the more significant early papers on metabolome analysis are listed in the references below.^{[6][7][8][9]}

Measuring the metabolome

The metabolome reflects the interaction between an organism's genome and its environment. As a result, an organism's metabolome can serve as an excellent probe of its phenotype (i.e. the product of its genotype and its environment). Metabolites can be measured (identified, quantified or classified) using a number of different technologies including NMR spectroscopy and mass spectrometry. Most mass spectrometry (MS) methods must be coupled to various forms of liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis (CE) to facilitate compound separation. Each method is typically able to identify or characterize 50-5000 different metabolites or metabolite "features" at a time, depending on the instrument or protocol being used. Currently it is not possible to analyze the entire range of metabolites by a single analytical method.

Nuclear magnetic resonance (NMR) spectroscopy is an analytical chemistry technique that measures the absorption of radiofrequency radiation of specific nuclei when molecules containing those nuclei are placed in strong magnetic fields. The frequency (i.e. the chemical shift) at which a given atom or nucleus absorbs is highly dependent on the chemical environment (bonding, chemical structure nearest neighbours, solvent) of that atom in a given molecule. The NMR absorption patterns produce "resonance" peaks at different frequencies or different chemical shifts – this collection of peaks is called an NMR spectrum. Because each chemical compound has a different chemical structure, each compound will have a unique (or almost unique) NMR spectrum. As a result, NMR is particularly useful for the characterization, identification and quantification of small molecules, such as metabolites. The widespread use of NMR for "classical" metabolic studies, along with its exceptional capacity to handle complex metabolite mixtures is likely the reason why NMR was one of the first technologies to be widely adopted for routine metabolome measurements. As an analytical technique, NMR is non-destructive, non-biased, easily quantifiable, requires little or no separation, permits the identification of novel compounds and it needs no chemical derivatization. NMR is particularly amenable to detecting compounds that are less tractable to LC-MS analysis, such as sugars, amines or volatile liquids or GC-MS analysis, such as large molecules (>500 Da) or relatively non-reactive compounds. NMR is not a very sensitive technique with a lower limit of detection of about 5 μ M. Typically 50-150 compounds can be identified by NMR-based metabolomic studies.

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio of molecules. Molecules or molecular fragments are typically charged or ionized by spraying them through a charged field (electrospray ionization), bombarding them with electrons from a hot filament (electron ionization) or blasting them with a laser when they are placed on specially coated plates (matrix assisted laser desorption ionization). The charged molecules are then propelled through space using electrodes or magnets and their speed, rate of curvature, or other physical characteristics are measured to determine their mass-to-charge ratio. From these data the mass of the parent molecule can be determined. Further fragmentation of the molecule through controlled collisions with gas molecules or with electrons can help determine the structure of molecules. Very accurate mass measurements can also be used to determine the elemental formulas or elemental composition of compounds. Most forms of mass spectrometry require some form of separation using liquid chromatography or gas chromatography. This separation step is required to simplify the resulting mass spectra and to permit more accurate compound identification. Some mass spectrometry methods also require that the molecules be derivatized or chemically modified so that they are more

amenable for chromatographic separation (this is particularly true for GC-MS). As an analytical technique, MS is a very sensitive method that requires very little sample (<1 ng of material or <10 µL of a biofluid) and can generate signals for thousands of metabolites from a single sample. MS instruments can also be configured for very high throughput metabolome analyses (hundreds to thousands of samples a day). Quantification of metabolites and the characterization of novel compound structures is more difficult by MS than by NMR. LC-MS is particularly amenable to detecting hydrophobic molecules (lipids, fatty acids) and peptides while GC-MS is best for detecting small molecules (<500 Da) and highly volatile compounds (esters, amines, ketones, alkanes, thiols).

Unlike the genome or even the proteome, the metabolome is a highly dynamic entity that can change dramatically, over a period of just seconds or minutes. As a result, there is growing interest in measuring metabolites over multiple time periods or over short time intervals using modified versions of NMR or MS-based metabolomics.

Metabolome databases

Because an organism's metabolome is largely defined by its genome, different species will have different metabolomes. Indeed, the fact that the metabolome of a tomato is different than the metabolome of an apple is the reason why these two fruits taste so different. Furthermore, different tissues, different organs and biofluids associated with those organs and tissues can also have distinctly different metabolomes. The fact that different organisms and different tissues/biofluids have such different metabolomes has led to the development of a number of organism-specific and biofluid-specific metabolome databases. Some of the better known metabolome databases include the Human Metabolome Database or HMDB,^[10] the Yeast Metabolome Database or YMDB,^[11] the E. coli Metabolome Database or ECMDB,^[12] the Arabidopsis metabolome database or AraCyc^[13] as well as the Urine Metabolome Database,^[14] the Cerebrospinal Fluid (CSF) Metabolome Database^[15] and the Serum Metabolome Database.^[16] The latter three databases are specific to human biofluids. A number of very popular general metabolite databases also exist including KEGG,^[17] MetaboLights,^[18] the Golm Metabolome Database,^[19] MetaCyc,^[20] LipidMaps^[21] and Metlin.^[22] Metabolome databases can be distinguished from metabolite databases in that metabolite databases contain lightly annotated or synoptic metabolite data from multiple organisms while metabolome databases contain richly detailed and heavily referenced chemical, pathway, spectral and metabolite concentration data for specific organisms.

The Human Metabolome Database

The Human Metabolome Database is a freely available, open-access database containing detailed data on more than 40,000 metabolites that have already been identified or are likely to be found in the human body. The HMDB contains three kinds of information: 1) chemical information, 2) clinical information, and 3) biochemical information. The chemical data includes >40,000 metabolite structures with detailed descriptions, extensive chemical classifications, synthesis information and observed/calculated chemical properties. It also contains nearly 10,000 experimentally measured NMR, GC-MS and LC/MS spectra from more than 1100 different metabolites. The clinical information includes data on >10,000 metabolite-biofluid concentrations, metabolite concentration information on more than 600 different human diseases and pathway data for more than 200 different inborn errors of metabolism. The biochemical information includes nearly 6000 protein (and DNA) sequences and more than 5000 biochemical reactions that are linked to these metabolite entries. The HMDB supports a wide variety of online queries including text searches, chemical structure searches, sequence similarity searches and spectral similarity searches. This makes it particularly useful for metabolomic researchers who are attempting to identify or understand metabolites in clinical metabolomic studies. The first version of the HMDB was released in Jan. 1 2007 and was compiled by scientists at the University of Alberta and the University of Calgary. At that time they reported data on 2,500 metabolites, 1,200 drugs and 3,500 food components. Since

then these scientists have greatly expanded the collection. The latest version of the HMDB (version 3.5) contains >16,000 endogenous metabolites, >1500 drugs and >22,000 food constituents or food metabolites.^[23]

Human biofluid metabolomes

Scientists at the University of Alberta have been systematically characterizing specific biofluid metabolomes including the serum metabolome,^[16] the urine metabolome,^[14] the cerebrospinal fluid (CSF) metabolome^[15] and the saliva metabolome. These efforts have involved both experimental metabolomic analysis (involving NMR, GC-MS, ICP-MS, LC-MS and HPLC assays) as well as extensive literature mining. According to their data, the human serum metabolome contains at least 4200 different compounds (including many lipids), the human urine metabolome contains at least 3000 different compounds (including hundreds of volatiles and gut microbial metabolites), the human CSF metabolome contains nearly 500 different compounds while the human saliva metabolome contains approximately 400 different metabolites, including many bacterial products.

Yeast Metabolome Database

The Yeast Metabolome Database is a freely accessible, online database of >2000 small molecule metabolites found in or produced by Saccharomyces cerevisiae (Baker's yeast). The YMDB contains two kinds of information: 1) chemical information and 2) biochemical information. The chemical information in YMDB includes 2027 metabolite structures with detailed metabolite descriptions, extensive chemical classifications, synthesis information and observed/calculated chemical properties. It also contains nearly 4000 NMR, GC-MS and LC/MS spectra obtained from more than 500 different metabolites. The biochemical information in YMDB includes >1100 protein (and DNA) sequences and >900 biochemical reactions. The YMDB supports a wide variety of queries including text searches, chemical structure searches, sequence similarity searches and spectral similarity searches. This makes it particularly useful for metabolomic researchers who are studying yeast as a model organism or who are looking into optimizing the production of fermented beverages (wine, beer).

Secondary electrospray ionization-high resolution mass spectrometry SESI-HRMS is a non-invasive analytical technique that allows us to monitor the yeast metabolic activities. SESI-HRMS has found around 300 metabolites in the yeast fermentation process, this suggests that a large number of glucose metabolites are not reported in the literature.^[24]

The Escherichia coli Metabolome Database

The E. Coli Metabolome Database is a freely accessible, online database of >2700 small molecule metabolites found in or produced by Escherichia coli (E. coli strain K12, MG1655). The ECMDB contains two kinds of information: 1) chemical information and 2) biochemical information. The chemical information includes more than 2700 metabolite structures with detailed metabolite descriptions, extensive chemical classifications, synthesis information and observed/calculated chemical properties. It also contains nearly 5000 NMR, GC-MS and LC-MS spectra from more than 600 different metabolites. The biochemical information includes >1600 protein (and DNA) sequences and >3100 biochemical reactions that are linked to these metabolite entries. The ECMDB supports many different types of online queries including text searches, chemical structure searches, sequence similarity searches and spectral similarity searches. This makes it particularly useful for metabolomic researchers who are studying E. coli as a model organism.

Secondary electrospray ionization (SESI-MS) can discriminate between eleven E. Coli strains thanks to the volatile organic compound profiling.^[25]

See also

- [Tumor metabolome](#)
- [Protein electrophoresis](#)
- [Protein sequencing](#)

References

1. Oliver SG, Winson MK, Kell DB, Baganz F (September 1998). "Systematic functional analysis of the yeast genome". *Trends in Biotechnology*. **16** (9): 373–8. [CiteSeerX 10.1.1.33.5221](#) (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.5221>). doi:10.1016/S0167-7799(98)01214-1 (<https://doi.org/10.1016%2FS0167-7799%2898%2901214-1>). PMID 9744112 (<https://pubmed.ncbi.nlm.nih.gov/9744112>).
2. Wishart DS (September 2007). "Current progress in computational metabolomics" (<https://doi.org/10.1093%2Fbib%2Fbbm030>). *Briefings in Bioinformatics*. **8** (5): 279–93. doi:10.1093/bib/bbm030 (<https://doi.org/10.1093%2Fbib%2Fbbm030>). PMID 17626065 (<https://pubmed.ncbi.nlm.nih.gov/17626065>).
3. Nordström A, O'Maille G, Qin C, Siuzdak G (May 2006). "Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705959>). *Analytical Chemistry*. **78** (10): 3289–95. doi:10.1021/ac060245f (<https://doi.org/10.1021%2Fac060245f>). PMC 3705959 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705959>). PMID 16689529 (<https://pubmed.ncbi.nlm.nih.gov/16689529>).
4. Tweeddale H, Notley-McRobb L, Ferenci T (October 1998). "Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107546>). *Journal of Bacteriology*. **180** (19): 5109–16. doi:10.1128/JB.180.19.5109-5116.1998 (<https://doi.org/10.1128%2FJB.180.19.5109-5116.1998>). PMC 107546 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107546>). PMID 9748443 (<https://pubmed.ncbi.nlm.nih.gov/9748443>).
5. Harrigan GG, Goodacre R, eds. (2003). *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis*. Boston: Kluwer Academic Publishers. ISBN 978-1-4020-7370-0.
6. Fiehn O, Kloska S, Altmann T (February 2001). "Integrated studies on plant biology using multiparallel techniques". *Current Opinion in Biotechnology*. **12** (1): 82–6. doi:10.1016/S0958-1669(00)00165-8 (<https://doi.org/10.1016%2FS0958-1669%2800%2900165-8>). PMID 11167078 (<https://pubmed.ncbi.nlm.nih.gov/11167078>).
7. Fiehn O (2001). "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447208>). *Comparative and Functional Genomics*. **2** (3): 155–68. doi:10.1002/cfg.82 (<https://doi.org/10.1002%2Fcfg.82>). PMC 2447208 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447208>). PMID 18628911 (<https://pubmed.ncbi.nlm.nih.gov/18628911>).
8. Weckwerth W (2003). "Metabolomics in systems biology" (<https://semanticscholar.org/paper/bd7eb46b953c9b97ec64f7efe1fa05483751a3b3>). *Annual Review of Plant Biology*. **54**: 669–89. doi:10.1146/annurev.arplant.54.031902.135014 (<https://doi.org/10.1146%2Fannurev.arplant.54.031902.135014>). PMID 14503007 (<https://pubmed.ncbi.nlm.nih.gov/14503007>). S2CID 1197884 (<https://api.semanticscholar.org/CorpusID:1197884>).
9. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (May 2004). "Metabolomics by numbers: acquiring and understanding global metabolite data". *Trends in Biotechnology*. **22** (5): 245–52. doi:10.1016/j.tibtech.2004.03.007 (<https://doi.org/10.1016%2Fj.tibtech.2004.03.007>). PMID 15109811 (<https://pubmed.ncbi.nlm.nih.gov/15109811>).
10. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. (January 2007). "HMDB: the Human Metabolome Database" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1899095>). *Nucleic Acids Research*. **35** (Database issue): D521-6. doi:10.1093/nar/gkl923 (<https://doi.org/10.1093%2Fnar%2Fgkl923>). PMC 1899095 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1899095>). PMID 17202168 (<https://pubmed.ncbi.nlm.nih.gov/17202168>).