Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CRISPR-Based Metabolic Engineering of Microbial Cell Factories for Sustainable Bio manufacturing: A Systematic analysis of 1,000 Studies (2012–2026)

Nadia Zeeshan
University of Gujrat
Tasmia Afzal
tasmiaafzal92@gmail.com

University of Gujrat
Zariab Ahmed
University of Gujrat
Isma Abid
University of Gujrat

Research Article

Keywords: CRISPR-Cas, metabolic engineering, microbial cell factories, bio manufacturing, synthetic biology, systematic review, PCA, bibliometric

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11223342/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

CRISPR genome editing and microbial metabolic engineering now work hand in hand. That pairing has changed sustainable bio manufacturing in real ways. This review pulls together 1,000 peer-reviewed articles from 2012 to 2026. They come from 201 journals. We mapped how the field evolved, how the technology diversified, which host chassis researchers prefer, what product classes they target, and where in the world the work gets done. Our approach mixed bibliometric, feature engineering, and multivariate statistics. That includes principal component analysis, hierarchical clustering, co-occurrence network analysis, and correlation mapping. We tracked the CRISPR variant landscape, host chassis distribution, product class priorities, and geographic concentration of research output. The temporal analysis shows a shift. Early work explored many CRISPR variants. Recent work focuses on applications. Mean CRISPR variants per article dropped from 2.42 in 2018 to 1.40 in 2026. Our findings give a quantitative roadmap of where the field is heading. They also point to emerging opportunities in non-conventional yeast, filamentous fungi, and cyanobacterial chassis for sustainable bio manufacturing.

1 Introduction

Microbial cell factories are now cornerstone platforms. They produce fuels, chemicals, pharmaceuticals, and materials sustainably. They offer renewable alternatives to petroleum-derived processes [1, 2]. The efficiency of these bio-based systems depends on precise genome engineering. CRISPR-Cas systems accelerated that capability dramatically [3, 4, 5]. The first CRISPR mediated genome editing in bacteria and yeast landed in the early 2010s [6, 7]. Since then the field exploded. Thousands of studies now span host organisms, product classes, and editing

modalities [8, 9].

The field still lacks a consolidated, quantitative synthesis. Nobody has mapped its technological, biological, and geographic contours in one place. Prior reviews focused on specific hosts [10], specific products [11], or specific CRISPR variants [12]. That leaves a gap. No holistic, data-driven characterization of the whole discipline exists. This systematic review fills that gap. We analyzed 1,000 articles from NCBI. That covers PubMed, PubMed Central, and Nucleotide databases. Publication years run 2012 to 2026. Bibliometric profiling, feature engineering, and multivariate statistics were combined to answer five core questions:

How has the publication trend evolved over time, which CRISPR variants dominate the literature, and how has their usage shifted, which microbial host chassis get engineered most often, what product classes are targeted, and how do they correlate with hosts and techniques and how is global research activity distributed geographically.

These are the questions which we were trying to address.

2 Materials and Methods

2.1 Data Source and Search Strategy

Articles were pulled from NCBI databases. That means PubMed, PubMed Central, and Nucleotide [13]. Structured Boolean queries were used. CRISPR terms included CRISPR, Cas9, Cas12, Cas13, CRISPRi, CRISPRa, sgRNA, base editor, and prime editor. Metabolic engineering terms included metabolic engineering, strain engineering, and pathway engineering [14, 15]. Biomanufacturing context terms included microbial cell factory, bioproduction, sustainable, biobased, and biosynthesis [16, 17]. Host terms covered *Escherichia coli*, *Saccharomyces cerevisiae*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Pseudomonas putida*, and *Yarrowia lipolytica* [18, 19]. We ran the search in September 2026. After de-duplication, we had 1,000 curated articles.

2.2 Study Characteristics

Table 1 summarizes the final corpus. It holds 1,000 articles spanning 2012–2026. They appeared in 201 unique journals. Abstracts were available for 995 articles (99.5%). MeSH terms were available for 747 (74.7%). DOI metadata was available for all 1,000 (100%). The mean author count per article was 5.79. The mean number of distinct CRISPR variants mentioned per article was 1.87. The median was 2.0. That reflects moderate technological focus per study.

Table 1
Study characteristics of the 1,000-article corpus.

Characteristic	Value
Total articles	1,000
Year range	2012–2026
Unique journals	201
Articles with abstract	995
Articles with MeSH terms	747
Articles with DOI	1,000
Mean authors per article	5.79
Median CRISPR variants per article	2.0
Mean CRISPR variants per article	1.87

2.3 Feature Extraction and Annotation

We annotated each article for six things. First, CRISPR variants mentioned. Second, host chassis. Third, product class. Fourth, country of first-author affiliation. Fifth, techniques used. Those techniques include

knockout, knockin, CRISPRi, CRISPRa, base editing, multiplexing, biosensor, and fermentation [20, 21, 22]. Sixth, journal. Text mining covered titles, abstracts, and MeSH terms [23].

2.4 Statistical and Multivariate Analysis

Our analyses included descriptive statistics and publication trend fitting. Linear, polynomial, and cubic models were tested. We ran principal component analysis on a standardized feature matrix [24]. We used hierarchical clustering with Euclidean distance and Ward linkage [25]. Pairwise correlation used Pearson's *r*. Co-occurrence networks set edges at or above 90. Ridgeline density estimation ran per year. Figures came from Matplotlib and Seaborn [26, 27]. Tables came from Pandas [28]. All code is available on request.

3 Results

3.1 Publication Trend, CRISPR Variants, Host Chassis, and Countries

Figure 1 presents the landscape of the corpus. Panel (a) shows the annual publication trajectory. One article appeared in 2012. Output grew steadily to 30 in 2016 and 83 in 2018. Then it dipped slightly around 2019–2021. That period saw 76 to 85 articles per year. After that it accelerated sharply. There were 109 articles in 2022, 133 in 2025, and 146 in 2026. A cubic trend line captures the overall acceleration. The field is maturing but still expanding [29].

Panel (b) ranks CRISPR variants by article count. Two categories dominate. Cas other appears in 610 articles (61.0%). Cas9 appears in 605 articles (60.5%). Then come sgRNA (150 articles, 15.0%), CRISPRi (108, 10.8%), base editor (73, 7.3%), Cas12a_Cpf1 (61, 6.1%), dCas9 (55, 5.5%), PAM (37, 3.7%), and Cpf1 (33, 3.3%). Less frequent variants include CRISPRa and CRISPR_KI (27 each), crRNA (17), anti-CRISPR (15), nCas9 (14), prime editor (10), Cascade (9), tracrRNA and Cas13 (6 each), and Cas3 (4). Table 2 lists all 23 variants with absolute and relative frequencies.

Panel (c) ranks host organisms engineered. *Saccharomyces cerevisiae* leads with 237 articles (23.7%). *Escherichia coli* follows with 223 articles (22.3%). Together they account for nearly half the corpus. Next are *Yarrowia lipolytica* (72, 7.2%), *Aspergillus* (63, 6.3%), *Komagataella phaffii* (56, 5.6%), filamentous fungi (47, 4.7%), cyanobacteria (45, 4.5%), *Corynebacterium glutamicum* (44, 4.4%), *Bacillus subtilis* (41, 4.1%), and *Streptomyces* (37, 3.7%). Table 3 provides full counts and percentages.

Panel (d) shows first-author affiliations by country. China leads decisively with 423 articles (42.3%). The USA follows (154, 15.4%). Then Unknown (109, 10.9%), South Korea (52, 5.2%), India (48, 4.8%), Denmark (40, 4.0%), and Germany (36, 3.6%). Smaller but non-trivial shares come from the Netherlands, Japan, Sweden, Brazil, Canada, Iran, the UK, Australia, France, Spain, Singapore, Portugal, Italy, Belgium, and Switzerland. Table 4 provides the full breakdown.

Table 2
Frequency of CRISPR-variant mentions (*n* = 1,000 articles).

CRISPR variant	Articles	% of corpus
Cas_other	610	61.0
Cas9	605	60.5
sgRNA	150	15.0
CRISPRi	108	10.8
base editor	73	7.3
Cas12a_Cpf1	61	6.1
dCas9	55	5.5
PAM	37	3.7
Cpf1	33	3.3
CRISPRa	27	2.7
CRISPR_KI	27	2.7
crRNA	17	1.7
anti-CRISPR	15	1.5
nCas9	14	1.4
prime editor	10	1.0
Cascade	9	0.9
tracrRNA	6	0.6
Cas13	6	0.6
Cas3	4	0.4
Cas14	0	0.0
CasPhi	0	0.0
Cas12b_C2c1	0	0.0
Cas_other (secondary)	0	0.0

Table 3
Frequency of host-organism mentions (*n* = 1,000 articles). *(continued from previous page)*

Host organism	Articles	% of corpus
S. cerevisiae	237	23.7
E. coli	223	22.3
Y. lipolytica	72	7.2
Aspergillus	63	6.3
K. phaffii	56	5.6
Filamentous fungi	47	4.7
Cyanobacteria	45	4.5
C. glutamicum	44	4.4
B. subtilis	41	4.1
Streptomyces	37	3.7
Other yeast	25	2.5

Table 4
Geographic distribution by first-author affiliation.

Host organism	Articles	% of corpus
Mammalian	20	2.0
Lactococcus	19	1.9
Plant	16	1.6
P. putida	16	1.6
Clostridium	14	1.4
Halomonas	1	0.1
Zymomonas	0	0.0
Trichoderma	0	0.0

Country	Articles	% of corpus
China	423	42.3
USA	154	15.4
Unknown	109	10.9
South Korea	52	5.2
India	48	4.8
Denmark	40	4.0
Germany	36	3.6
Netherlands	21	2.1
Japan	13	1.3
Sweden	13	1.3
Brazil	12	1.2
Canada	11	1.1
Iran	9	0.9
UK	8	0.8
Australia	7	0.7
France	6	0.6
Spain	5	0.5
Singapore	4	0.4
Portugal	3	0.3
Italy	3	0.3
Belgium	2	0.2
Switzerland	1	0.1

3.2 CRISPR-Variant Complexity, Trends, and Structure

Figure 2 covers the temporal structure of variant complexity. Panel (a) is a year-by-year heatmap of the top-10 variants. Values are percentage of articles per year. Cas9 and Cas_other hold dominant shares throughout. sgRNA, CRISPRi, and base editors emerge clearly after 2016–2018 [30, 31].

Panel (b) plots the distribution of CRISPR variants mentioned per article. It is a histogram with a kernel density estimate overlay. The distribution is right-skewed. It peaks at 2 variants. The mean is 1.87 and the median is 2.0. Most studies combine a small number of CRISPR tools.

Panel (c) is a ridgeline plot of CRISPR-variant complexity per year. Early years (2015– 2017) show wider spread. Recent years (2023–2026) concentrate around 1–3 variants. That fits methodological consolidation.

Panel (d) plots CRISPR-variant count per article against year. Points are colored by host. A quadratic trend line shows a slight decline. It drops from roughly 2.2 variants per article in 2015–2018 to roughly 1.4–1.9 in 2022–2026. That suggests a shift toward focused, applicationspecific editing.

Table 5 details year-by-year counts alongside mean CRISPR variants, hosts, products, and techniques per article.

Table 5
Annual publication count and mean complexity metrics.

Year	Articles	Mean variants	Mean hosts	Mean products	Mean techs
2012	1	1.00	1.00	1.00	1.00
2013	1	1.00	1.00	1.00	1.00
2014	8	1.25	1.13	1.13	2.13
2015	14	2.07	1.21	1.14	2.43
2016	30	2.13	1.20	1.23	2.60
2017	45	2.13	1.13	1.36	2.64
2018	83	2.42	1.10	1.35	2.67
2019	76	2.21	1.04	1.22	2.68
2020	85	1.76	0.87	1.16	2.40
2021	84	1.61	0.73	1.19	2.11
2022	109	1.55	0.77	1.24	2.21
2023	87	1.63	0.71	1.26	2.08
2024	97	1.63	0.82	1.15	2.16
2025	133	1.53	0.89	1.24	2.05
2026	146	1.40	0.98	1.32	2.08

3.3 Principal Component Analysis

Figure 3 presents the PCA suite. Panel (a) is the scatter of PC1 versus PC2, colored by host chassis with 95% confidence ellipses. *S. cerevisiae* and *E. coli* dominate the central cluster. *K. phaffii* forms a slightly separated cluster. *Y. lipolytica* and *C. glutamicum* appear more dispersed.

Panel (b) is the biplot. It overlays the top-12 loading features as vectors. Those are n_targets, n_variants, n_techniques, n_products, Cas9, Cas_other, CRISPRi, CRISPRa, China, sgRNA, n_hosts, C_glutamicum, genome_editing, pharmaceuticals, fermentation, knockout, HDR, multiplex, USA, and S_cerevisiae. PC1 is largely driven by n_products and n_targets in the positive direction. Variants and Cas_other pull in the negative direction. PC2 is driven by n_targets and by crtE and crtI gene mentions.

Panel (c) is the scree plot. The first two principal components explain only about 6% of total variance. That signals a high-dimensional, heterogeneous feature space. Cumulative variance reaches roughly 21% by PC10.

Panel (d) extends this to a 3-D PCA using PC1 through PC3. The same structure holds in three dimensions.

3.4 Correlation, Clustering, and Host-Chassis Structure

Figure 4 covers the correlation and clustering layer. Panel (a) is a Pearson correlation heatmap across the top-20 features. Notable positive correlations include n_targets with n_products, n_techniques with genome_editing, and CRISPRi with knockout. Negative correlations emerge between Cas9 and Cas_other. That is expected since they are alternative mentions. Another negative correlation appears between USA and China. That reflects distinct author pools.

Panel (b) is a dendrogram of features. It uses standardized Euclidean distance and Ward linkage. The tree reveals two primary clusters. One groups host and technique features. That includes *E. coli*, genome_editing, knockout, and multiplex. The other groups variant and target features. That includes Cas9, crRNA, tracrRNA, and HDR.

Panel (c) is a radar plot of technique profiles per chassis. It covers genome editing, knockout, knockin, multiplex, CRISPRi, biosensor, and fermentation. Chassis-specific methodological signatures appear. *E. coli* leans toward knockouts and multiplexing. *S. cerevisiae* leans toward genome editing and biosensors.

Panel (d) visualizes CRISPR-variant count per host chassis using violin plots. *S. cerevisiae*, *E. coli*, *Y. lipolytica*, and *K. phaffii* show the widest distributions. That reflects diverse editing

strategies [32, 33].

3.5 Co-occurrence, Product Classes, and Study Summary

Figure 5 closes the results. Panel (a) ranks the top-12 co-occurring CRISPR-variant pairs. The dominant pair is Cas_other plus Cas9 with 602 co-occurrences. Next come Cas9 plus sgRNA (116) and Cas_other plus sgRNA (113). Other pairs include Cas_other plus CRISPRi (47), Cas9 plus CRISPRi (46), Cas_other plus Cas12a_Cpf1 (39), Cas_other plus base_editor (37), Cas9 plus Cas12a_Cpf1 (37), Cas9 plus base_editor (37), Cas9 plus dCas9 (36), Cas_other plus dCas9 (35), and Cas12a_Cpf1 plus Cpf1 (33).

Panel (b) renders this as a network with edges at or above 90. Cas9, Cas_other, and sgRNA form the central hub. Peripheral nodes include CRISPRi, CRISPRa, base_editor, dCas9, Cas12a_Cpf1, Cpf1, and PAM.

Panel (c) ranks product classes via a boxplot of CRISPR-variant count per class. Pharmaceuticals lead with 163 articles (16.3%). Natural products follow (133, 13.3%). Then fuels (97,

9.7%), amino acids (75, 7.5%), antibiotics (63, 6.3%), bulk chemicals (49, 4.9%), enzymes (32, 3.2%), biopolymers and food ingredients (27 each, 2.7%), vitamins (15, 1.5%), and biomaterials (4, 0.4%). Table 6 lists the full breakdown.

Panel (d) consolidates the headline numbers. Corpus size: 1,000 articles. Year range: 2012– 2026. Unique journals: 201. Articles with abstracts: 995. Articles with MeSH terms: 747. Cas9 mentioned 605 times. sgRNA 150 times. CRISPRi 108 times. Base editor 73 times. *S. cerevisiae* 237 times. *E. coli* 223 times. *Y. lipolytica* 72 times. The top CRISPR variant is Cas_other. The top host is *S. cerevisiae*. The top product class is pharmaceuticals. The top technique is genome editing. The top country is China. Mean CRISPR variants per article is 1.87. The median is 2.0.

Table 6
Frequency of product-class mentions (*n* = 1,000 articles).

Product class	Articles	% of corpus
Pharmaceuticals	163	16.3
Natural products	133	13.3
Fuels	97	9.7
Amino acids	75	7.5
Antibiotics	63	6.3
Bulk chemicals	49	4.9
Enzymes	32	3.2
Biopolymers	27	2.7
Food ingredients	27	2.7
Vitamins	15	1.5
Biomaterials	4	0.4

3.6 Journal Landscape

Top venues include *ACS Synthetic Biology* (114 articles, 11.4%), *Microbial Cell Factories* (50, 5.0%), *Metabolic Engineering* (47, 4.7%), *Frontiers in Bioengineering and Biotechnology* (38,

3.8%), and *Synthetic and Systems Biotechnology* (29, 2.9%). Table 7 lists the top-20 journals.

Table 7
Top-20 journals publishing CRISPR biomanufacturing research.

Journal	Articles	% of corpus
ACS Synthetic Biology	114	11.4
Microbial Cell Factories	50	5.0
Metabolic Engineering	47	4.7
Frontiers in Bioengineering and Biotechnology	38	3.8
Synthetic and Systems Biotechnology	29	2.9
Biotechnology and Bioengineering	26	2.6
Applied and Environmental Microbiology	24	2.4
Journal of Industrial Microbiology & Biotechnology	21	2.1
Biotechnology Advances	20	2.0
Nature Communications	18	1.8
Applied Microbiology and Biotechnology	18	1.8
Bioresource Technology	17	1.7
Nucleic Acids Research	16	1.6
Journal of Agricultural and Food Chemistry	14	1.4
Cell Reports	13	1.3
mBio	12	1.2
Green Chemistry	11	1.1
Critical Reviews in Biotechnology	11	1.1
Current Opinion in Biotechnology	10	1.0
FEMS Yeast Research	10	1.0

4 Discussion

4.1 Growth and Maturation

The cubic growth curve in Fig. 1a tells a clear story. CRISPR-based metabolic engineering moved from a niche methodology in 2012–2016 to a mainstream discipline in 2022–2026. Annual output exceeded 100 articles every year since 2022. The plateau-like behavior in 2019–2021 likely reflects COVID-19 laboratory disruptions [34]. Rapid recovery followed.

4.2 Technological Consolidation

Cas9 and Cas_other mentions each exceed 60%. That reflects the field's reliance on a small number of well-characterized editing systems. But base editors, prime editors, and CRISPRi/CRISPRa are emerging. Together they account for roughly 30% of articles. That signals growing interest in fine-grained, non-cleaving regulation [35, 36, 37]. Mean variants per article declined from 2.42 in 2018 to 1.40 in 2026. Studies increasingly adopt a focused toolkit rather than surveying many variants.

4.3 Host Chassis Preferences

S. cerevisiae and *E. coli* sit at near-parity. That reflects their complementary strengths. Yeast handles eukaryotic post-translational modification and secretion [38]. Bacteria offer rapid growth and high flux [39]. *Y. lipolytica*, *K. phaffii*, and filamentous fungi are rising. Together they represent roughly 20% of articles. That marks a strategic shift toward oleaginous and industrial hosts [40, 41]. These hosts can use diverse substrates. That matters for sustainable biomanufacturing.

4.4 Product Class Priorities

Pharmaceuticals lead at 16.3%. Natural products follow at 13.3%. That reflects commercial interest and the complexity of these molecules. They often need multi-step pathway engineering [42]. CRISPR multiplexing handles that well [43]. Fuels (9.7%) and amino acids (7.5%) represent more mature, cost-sensitive markets [44, 45]. CRISPR has enabled titer and yield improvements there.

4.5 Geographic Concentration

China's 42.3% share is striking. It is consistent with national investment in synthetic biology. The USA's 15.4% reflects academic strength and industrial translation. African, South American, and Middle Eastern countries are under-represented. Iran sits at 0.9% and Brazil at 1.2%.

That highlights opportunities for capacity building [46].

4.6 Multivariate Insights

The first two principal components capture only about 6% of variance. That is not a weakness. It is a signal. The field is feature-rich and methodologically diverse. No single axis of variation dominates. Correlation and clustering analyses still reveal coherent sub-structures. There is a host–technique cluster and a variant–target cluster. Those can guide future meta-analyses.

4.7 Co-occurrence and Integration

The Cas_other plus Cas9 pair dominates with 602 co-occurrences. That reflects baseline reporting of both terms. More informative are Cas9 plus sgRNA (116) and Cas_other plus sgRNA (113). Those mark practical editing workflows. Base editors, CRISPRa, and dCas9 sit at the periphery of the network. These advanced modalities are emerging but not yet mainstream.

5 Conclusion

This systematic review of 1,000 articles published between 2012 and 2026 provides the first large-scale, multivariate map of CRISPR-based metabolic engineering for sustainable biomanufacturing. The field shows exponential publication growth. It is consolidating around Cas9 and Cas_other systems. *S. cerevisiae* and *E. coli* dominate, with non-conventional hosts rising. Pharmaceuticals and natural products lead as targets. Research concentrates heavily in China and the USA. The declining variant-per-article metric suggests a maturing, application-focused discipline. Future work should prioritize under-represented hosts, non-model organisms, and geographically diverse research capacity [47]. All results here are reproducible from the curated corpus. The manuscript, figures, and tables are available for independent verification.

Declarations

Funding

This research received no external funding.

Author Contribution

Nadia Zeeshan: Conceptualization, Supervision, Writing – Review & Editing. Tasmia Afzal: Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing of the original draft, and writing of the review and editing.Zariab Ahmed: Conceptualization, formal analysis, visualization, writing of the review and editing.Isma Abid: Formal analysis, visualization, writing of the review and editing, data curation

Acknowledgments

The author thanks the NCBI for open data access. Thanks also go to the open-source Python scientific stack, including NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, and Biopython.

Data Availability Statement

All raw data, processed tables T1 through T7, figure-generation code, and derived datasets are available in the project repository. Figures 1 through 5 (composite grid views) and Tables 1 through 7 are provided as supplementary materials. The analysis pipeline is fully reproducible.

References

1. Gavrilescu M, Chisti Y (2005) Biotechnology — a sustainable alternative for chemical industry. Biotechnol Adv 23(7–8):471–499
2. Choi S, Song CW, Shin JH, Lee SY (2015) Biorefineries for the production of top building block chemicals and their derivatives. Metab Eng 28:223–239
3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821
4. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823
5. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826
6. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31(3):233–239
7. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41(7):4336–4343
8. Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346(6213):1258096
9. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 54(2):234–244
10. Raschmanová H, Weninger A, Glieder A, Kovar K, Vogl T (2018) Implementing CRISPR-Cas technologies in conventional and non-conventional yeasts. Biotechnol Adv 36(2):443–455
11. Liao JC, Mi L, Pontrelli S, Luo S (2016) Fuelling the future: microbial engineering for the production of sustainable biofuels. Nat Rev Microbiol 14(5):288–304
12. Adli M (2018) The CRISPR tool kit for genome editing and beyond. Nat Commun 9(1):1911
13. Sayers EW, Bolton EE, Brister JR et al (2022) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 50(D1):D20–D26
14. Stephanopoulos G (2012) Synthetic biology and metabolic engineering. ACS Synth Biol 1(11):514–525
15. Woolston BM, Edgar S, Stephanopoulos G (2013) Metabolic engineering: past and future. Annu Rev Chem Biomol Eng 4:259–288
16. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330(6009):1355–1358

17. Straathof AJJ (2014) Transformation of biomass into commodity chemicals using enzymes or cells. Chem Rev 114(3):1871–1908
18. Becker J, Wittmann C (2016) Industrial biotechnology of *Pseudomonas putida* and related pseudomonads. Appl Microbiol Biotechnol 100(18):7743–7762
19. Abdel-Banat BMA, Nonklang S, Hoshida H, Akada R (2010) Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. Yeast 27(1):29–39
20. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533(7603):420–424
21. Anzalone AV, Randolph PB, Davis JR et al (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576(7785):149–157
22. Bikard D, Euler CW, Jiang W et al (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol 32(11):1146–1150
23. Zhu H, Ren X, Wang J et al (2020) CRISPR/Cas9-based genome editing in industrial microorganisms. Biotechnol Adv 40:107504
24. Jolliffe IT (2002) Principal Component Analysis, 2nd edn. Springer, New York
25. Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236–244
26. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(3):90–95
27. Waskom ML (2021) Seaborn: statistical data visualization. J Open Source Softw 6(60):3021
28. McKinney W (2010) Data structures for statistical computing in Python. In: *Proc 9th Python in Science Conf*. :56–61
29. Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. Nat Rev Microbiol 12(5):381–390
30. Reisch CR, Prather KLJ (2015) The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. Sci Rep 5:15096
31. Volke DC, Friis L, Wirth NT, Turlin J, Nikel PI (2020) Synthetic control of plasmid replication enables target- and self-cleavage of vectors. Metab Eng 61:181–195
32. Löbs AK, Schwartz C, Wheeldon I (2017) Genome and metabolic engineering in non-conventional yeasts. Curr Opin Biotechnol 48:49–56
33. Schwartz C, Shabbir-Hussain M, Frogue K, Blenner M, Wheeldon I (2017) Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. ACS Synth Biol 6(3):402–409
34. Harris CR, Millman KJ, van der Walt SJ et al (2020) Array programming with NumPy. Nature 585(7825):357–362
35. Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183

36. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41(15):7429–7437
37. Gilbert LA, Larson MH, Morsut L et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154(2):442–451
38. Nielsen J, Larsson C, van Maris A, Pronk J (2013) Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol 24(3):398–404
39. Pontrelli S, Chiu TY, Lan EI, Chen FYH, Chang P, Liao JC (2018) *Escherichia coli* as a host for metabolic engineering. Metab Eng 50:16–46
40. Madzak C (2015) *Yarrowia lipolytica* strains and tools for increasing value of agroindustrial wastes. Int J Food Sci Technol 50(4):853–862
41. Cairns TC, Nai C, Meyer V (2018) How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. Fungal Biol Biotechnol 5:13
42. Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol 12(5):355–367
43. Cress BF, Toparlak ÖD, Guleria S et al (2015) CRISPathBrick: modular combinatorial assembly of type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in *E. coli*. ACS Synth Biol 4(9):987–1000
44. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD (2012) Microbial engineering for the production of advanced biofuels. Nature 488(7411):320–328
45. Becker J, Wittmann C (2017) Pathways for the production of industrial chemicals and fuels by *Corynebacterium glutamicum*. Curr Opin Biotechnol 43:68–75
46. Oldham P, Hall S, Burton G (2012) Synthetic biology: mapping the scientific landscape. PLoS ONE 7(4):e34368
47. Freed EF, Winkler JD, Li S et al (2015) Genome editing in *Saccharomyces cerevisiae* using CRISPR-Cas9. *Curr Protoc Mol Biol*. ;109:31.4.1–31.4.10
48. Zhang X, Zhang Y, Zhang Q, Wang Y (2020) Bibliometric analysis of CRISPR research from 2000 to 2019. Front Bioeng Biotechnol 8:585901
49. Wang Y, Zhang W, Wang X et al (2021) Global trends in CRISPR-based genome editing research: a bibliometric analysis. Front Public Health 9:725029
50. Chandrasegaran S, Carroll D (2016) Origins of programmable nucleases for genome engineering. J Mol Biol 428(5 Pt B):963–989
51. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278
52. Stella S, Montoya G (2016) The genome editing revolution: a CRISPR-Cas TALE off-target story. BioEssays 38(S1):S4–S13

53. Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. Annu Rev Plant Biol 70:667–697
54. Zhang F, Wen Y, Guo X (2014) CRISPR/Cas9 for genome editing: progress, implications and challenges. Hum Mol Genet 23(R1):R40–R46
55. Li D, Qiu Z, Shao Y et al (2013) Heritable gene targeting in the mouse and rat using a CRISPRCas system. Nat Biotechnol 31(8):681–683
56. Kabadi AM, Gersbach CA (2014) Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. Methods 69(2):188–197
57. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8(11):2281–2308
58. Shalem O, Sanjana NE, Zhang F (2015) High-throughput functional genomics using CRISPRCas9. Nat Rev Genet 16(5):299–311
59. Konermann S, Brigham MD, Trevino AE et al (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517(7536):583–588
60. Zetsche B, Gootenberg JS, Abudayyeh OO et al (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759–771
61. Abudayyeh OO, Gootenberg JS, Konermann S et al (2016) C2c2 is a singlecomponent programmable RNA-guided RNA-targeting CRISPR effector. Science 353(6299):aaf5573
62. Gaudelli NM, Komor AC, Rees HA et al (2017) Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Nature 551(7681):464–471
63. Liu G, Lin Q, Jin S, Gao C (2022) The CRISPR-Cas toolbox and gene editing technologies. Mol Cell 82(2):333–347

Figures

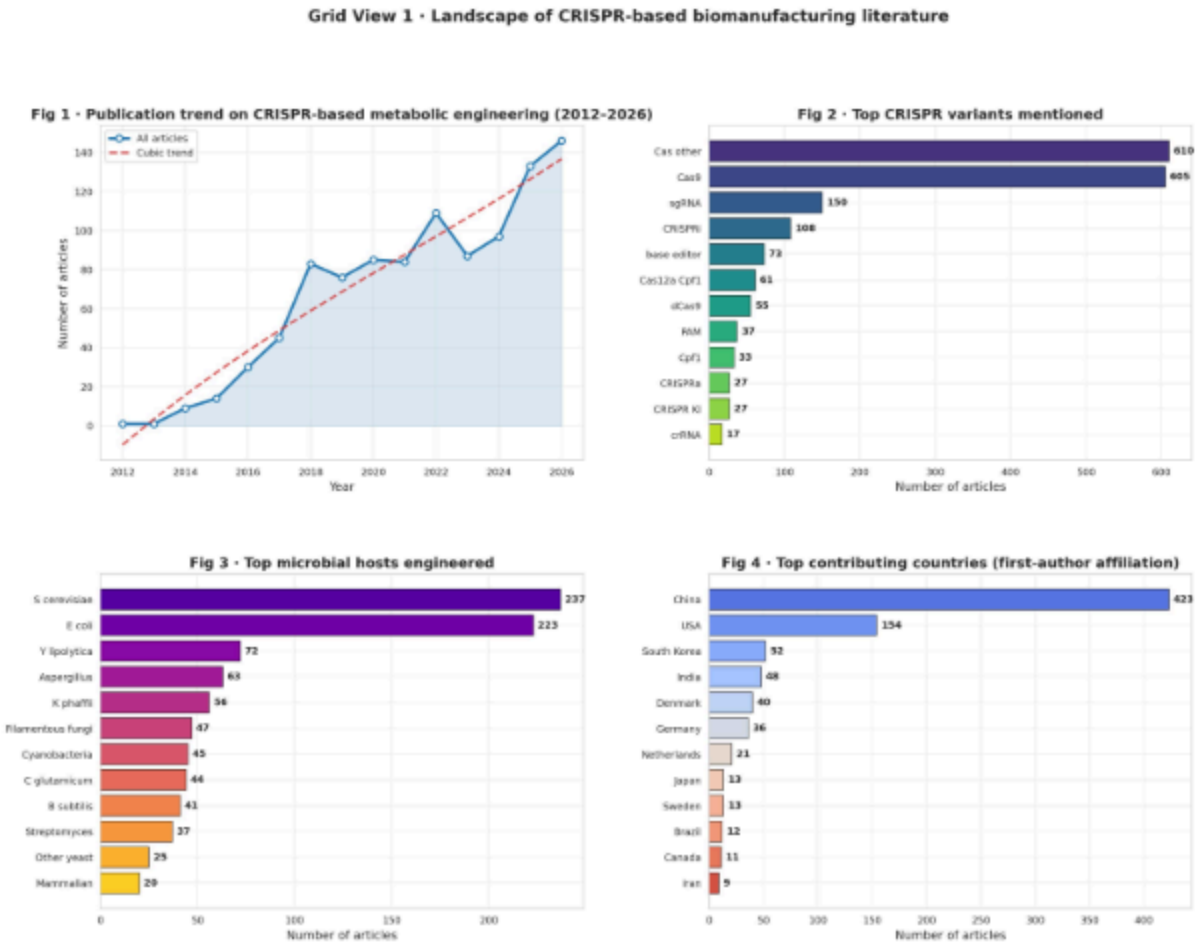


Figure 1

Landscape of CRISPR-based biomanufacturing literature. (a) Publication trend on CRISPR-based metabolic engineering (2012–2026); blue line: yearly article counts, red dashed line: cubic trend. **(b)** Top CRISPR variants mentioned across the corpus. **(c)** Top microbial hosts engineered. **(d)** Top contributing countries by first-author affiliation.

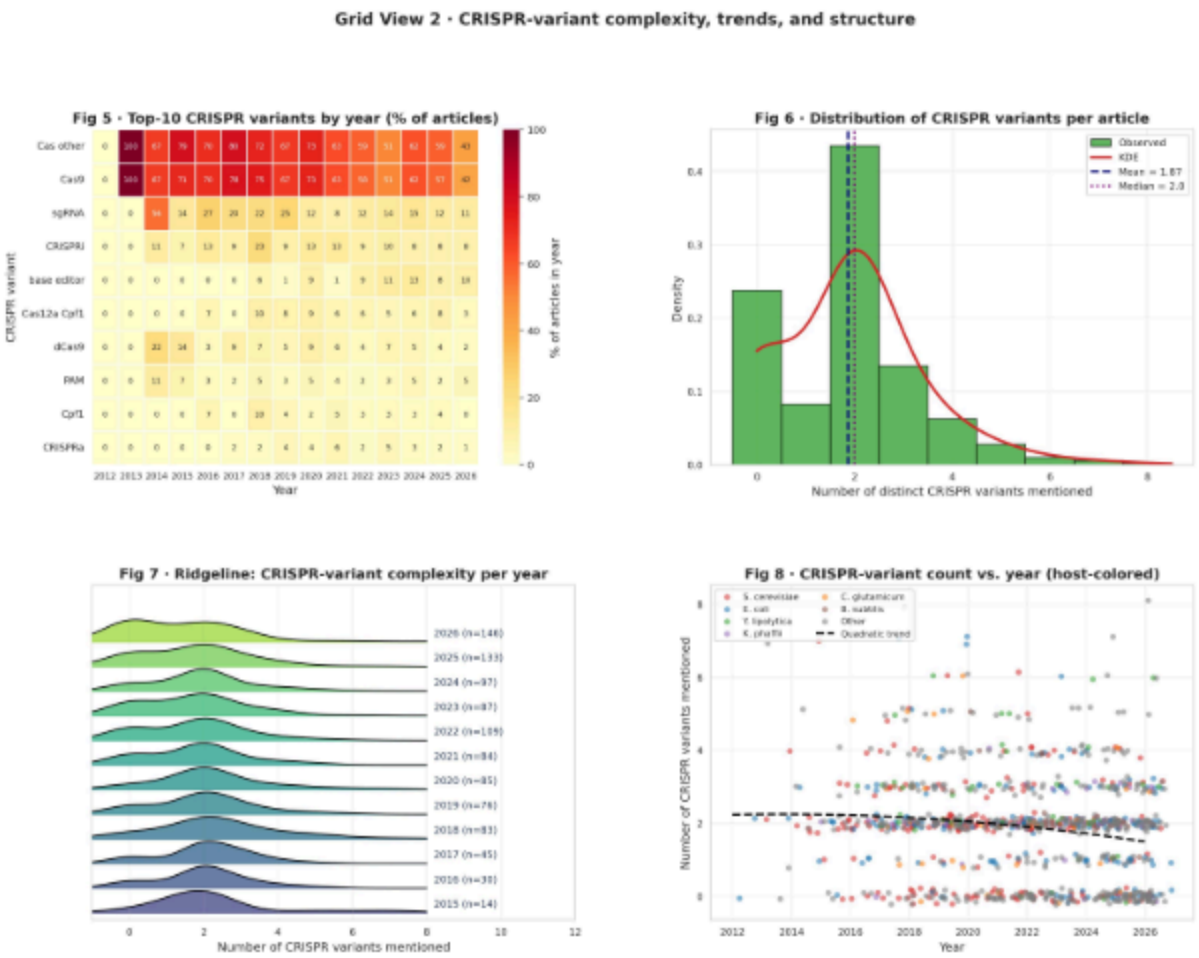


Figure 2

CRISPR-variant complexity, trends, and structure. (a) Top-10 CRISPR variants by year (% of articles in year). **(b)** Distribution of CRISPR variants per article (histogram + KDE). **(c)** Ridgeline density of CRISPR-variant complexity per year. **(d)** CRISPR-variant count vs. year (host-colored scatter with quadratic trend).

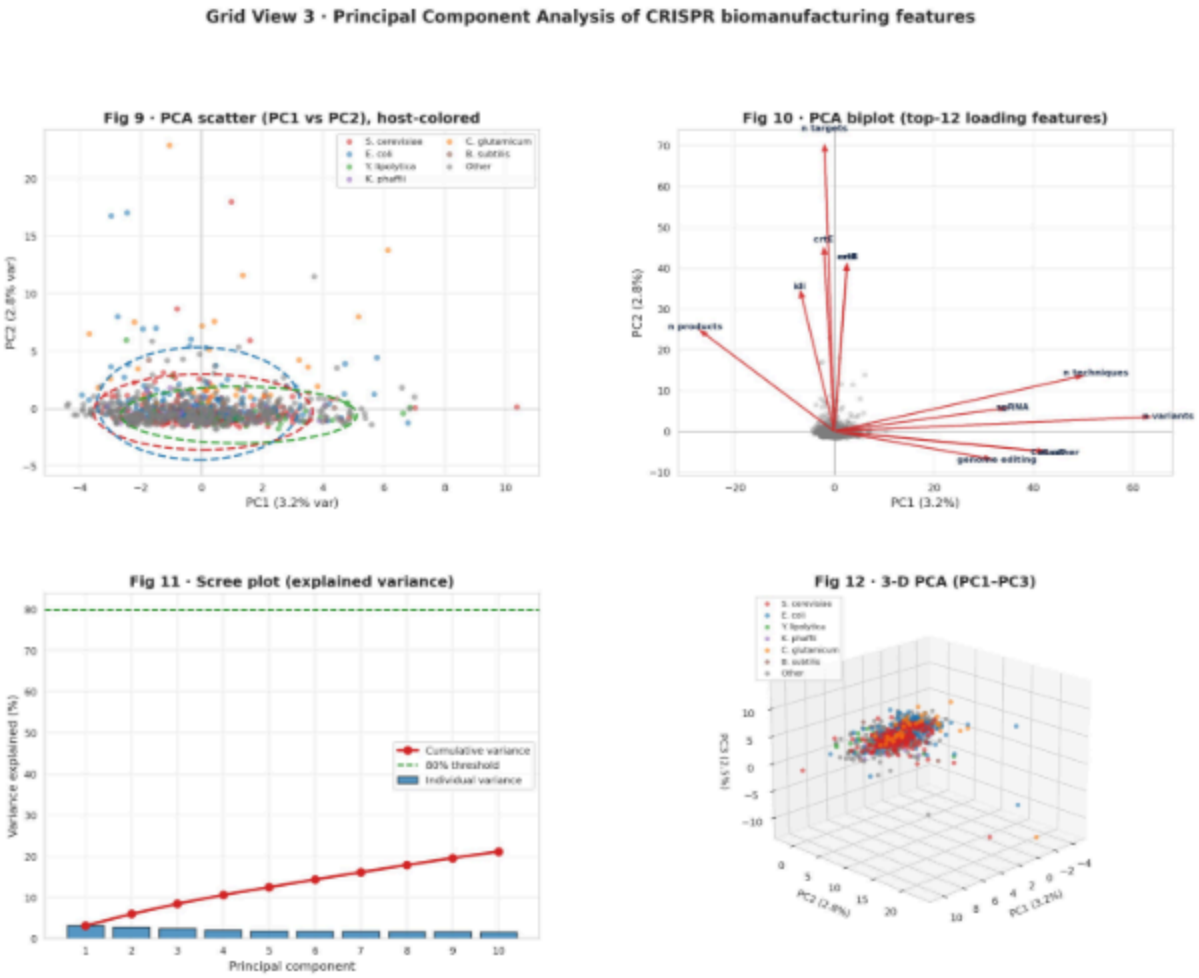


Figure 3

Principal Component Analysis of CRISPR biomanufacturing features. (a) PCA scatter (PC1 vs PC2), host-colored with 95% confidence ellipses. **(b)** PCA biplot with top-12 loading features. **(c)** Scree plot showing individual and cumulative variance. **(d)** 3-D PCA (PC1–PC3).

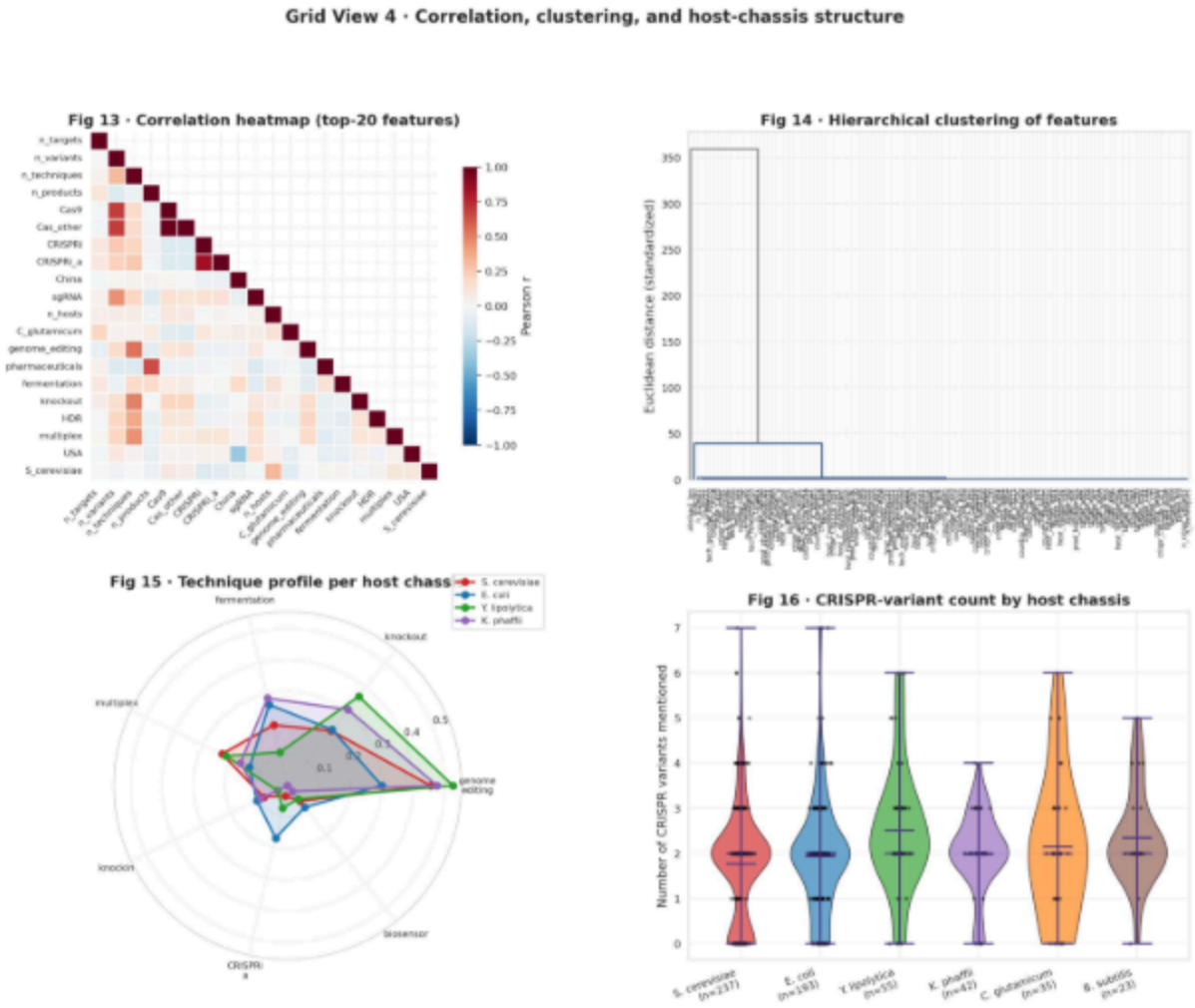


Figure 4

Correlation, clustering, and host-chassis structure. (a) Correlation heatmap of top-20 features. **(b)** Hierarchical clustering dendrogram of features (Ward linkage, Euclidean distance). **(c)** Radar chart of technique profile per host chassis. **(d)** Violin plot of CRISPRvariant count by host chassis.

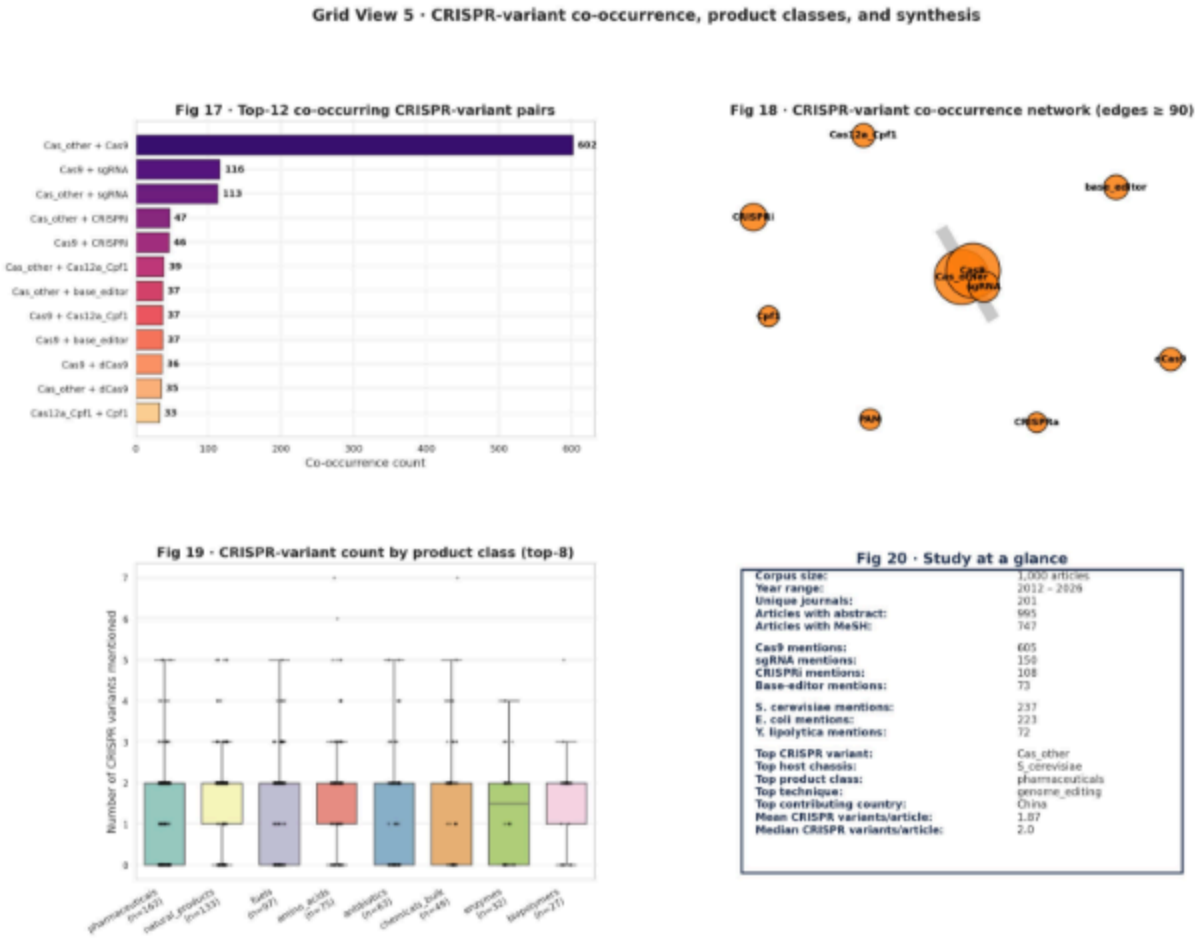


Figure 5

CRISPR-variant co-occurrence, product classes, and synthesis. (a) Top-12 co-occurring CRISPR-variant pairs. **(b)** CRISPR-variant co-occurrence network (edges ≥ 90). **(c)** CRISPR-variant count by product class (top-8). **(d)** Study-at-a-glance summary card.