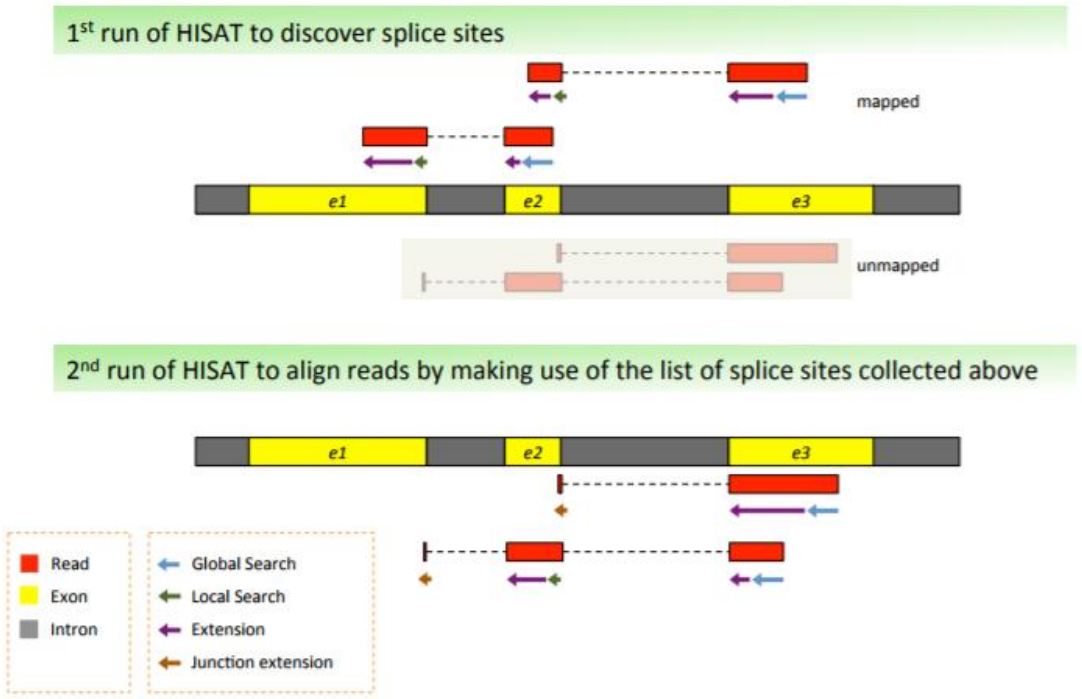


Figure S1 HISAT2 alignment principle

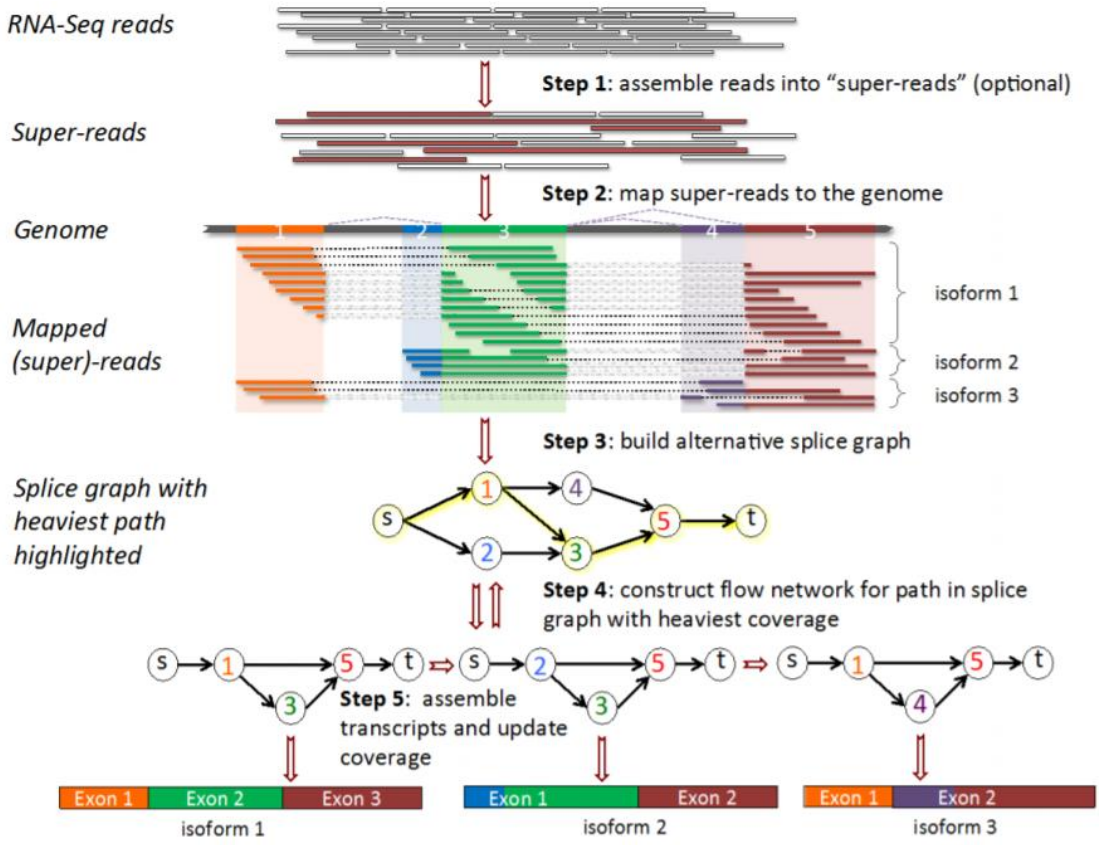


Figure S2 The principle of transcript assembly for Stringtie

Table S1 Gene differential expression analysis method

Type	Software	Standardized methods	PValue calculation model	FDR calculation method	Differential gene screening criteria
Biological duplication	DESeq2(Anders et al, 2014)	DESeq	Negative binomial distribution	BH	|log2(FC)| > 1,FDR< 0.05
No biological duplication	edgeR(Robinson et al, 2010)	TMM	Negative binomial distribution	BH	|log2(FC)| > 1,FDR< 0.05

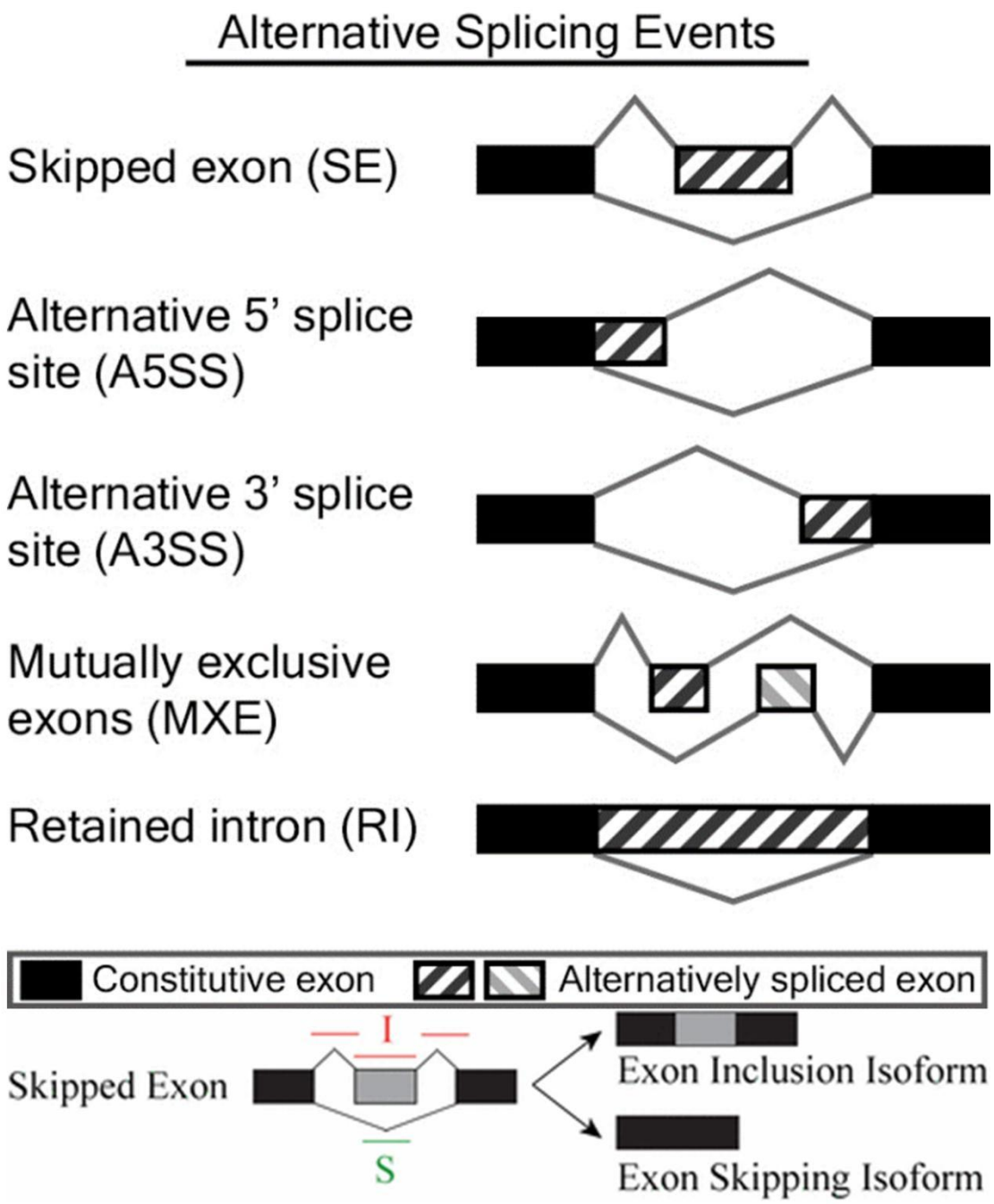


Figure S3 Schematic diagram of alternative splicing events

Code 1 Cross column duplicate filtering

```
import pandas as pd
from collections import defaultdict

# Read Excel file (assuming file name is input. xlsx)
df = pd.read_excel("input.xlsx", header=0)

# Create a mapping dictionary from values to column labels (automatic deduplication)
col_map = defaultdict(set)
```

```
for col in df.columns:
    for val in df[col].dropna().astype(str).unique():
        col_map[val].add(col)

# Create a mapping dictionary from column sets to values (automatic grouping)
group_map = defaultdict(set)
for val, cols in col_map.items():
    if len(cols) >= 2:
        key = ','.join(sorted(cols))    # Sort column labels alphabetically
        group_map[key].add(val)

# Generate the final result
result = [[cols, ','.join(sorted(values))]
          for cols, values in group_map.items()]

# Export to a new Excel file
pd.DataFrame(result, columns=["Comparison group", "Common
items"]).to_excel("output.xlsx", index=False)
```

Code 2 Q value search on demand

```
import pandas as pd

# Read Excel file
df = pd.read_excel("input.xlsx")

# Create a dictionary (with keys in the first column and values in the second column)
key_value = df.set_index(df.columns[0])[df.columns[1]].to_dict()

# Process the third column query and generate results
result = df[df.columns[2]].apply(
    lambda x: ",".join([str(key_value.get(k.strip(), "")) for k in str(x).split(",")])
)

# Save to a new Excel file
result.to_excel("output.xlsx", index=False, header=["Qvalue"])
```

Code 3 Merge with 3 decimal places retained

```
import pandas as pd

def format_row(go_str, val_str):
    # Force conversion to string processing
    go_ids = str(go_str).split(',')
    values = [float(v) for v in str(val_str).split(',')]
```

```
        # Process Scientific notation and format (use English brackets)
        formatted = []
        for g, v in zip(go_ids, values):
            # Change the parentheses here to English format
            formatted.append(f"{g.strip()}({v:.3f})")

        return ','.join(formatted)

# Force all columns to be of string type when reading
df = pd.read_excel('input.xlsx', dtype=str)

# Process each row of data
df['combined'] = df.apply(
    lambda row: format_row(row.iloc[0], row.iloc[1]),
    axis=1
)

# Save the Results
df[['combined']].to_excel('output.xlsx', index=False)
```

Code 4 Delete duplicate values

```
import pandas as pd

def merge_columns(input_path, output_path):
    # Read data and retain header rows
    df = pd.read_excel(input_path, header=0)
    headers = df.columns.tolist()    # Save original title

    # Processing data section (excluding header rows)
    df_data = df.map(lambda x: str(x).strip().replace(" ", "") if pd.notna(x) else "")

    # Cell level deduplication (retaining first occurrence)
    seen = set()
    # Traverse the indexes (rows, column names) of all non empty cells
    for (row_idx, col_name) in df_data.stack().index:
        # Convert column names to integer indexes
        col_idx = df_data.columns.get_loc(col_name)
        val = df_data.iat[row_idx, col_idx]
        if val:
            if val in seen:
                df_data.iat[row_idx, col_idx] = ""
            else:
                seen.add(val)
```

```
    # Extract valid data from each column and merge them
    merged_data = []
    # Traverse column index (integer)
    for col_idx in range(df_data.shape[1]):
        valid_values = [
            df_data.iat[row_idx, col_idx]
            for row_idx in df_data.index
            if df_data.iat[row_idx, col_idx]
        ]
        merged_data.append(", ".join(valid_values))

    # Build result (title row+merged data row)
    result_df = pd.DataFrame([headers, merged_data])

    # Save to Excel
    result_df.to_excel(output_path, index=False, header=False)

# Execute processing
merge_columns("input.xlsx", "output.xlsx")
```

Table S2 Data filtering statistics table

Sample	RawDatas	CleanData(%)	Adapter(%)	LowQuality(%)	polyA(%)	N(%)
H167-1	41731154	41597026 (99.68%)	20384 (0.05%)	113726 (0.27%)	0 (0.00%)	18 (0.00%)
H167-2	41707188	41500934 (99.51%)	71170 (0.17%)	135026 (0.32%)	0 (0.00%)	58 (0.00%)
H167-3	37465154	37334922 (99.65%)	21768 (0.06%)	108344 (0.29%)	0 (0.00%)	120 (0.00%)
H74-1	68290544	67974512 (99.54%)	39216 (0.06%)	276596 (0.41%)	0 (0.00%)	220 (0.00%)
H74-2	48068254	47833516 (99.51%)	63468 (0.13%)	171110 (0.36%)	0 (0.00%)	160 (0.00%)
H74-3	71589568	71355148 (99.67%)	36824 (0.05%)	197576 (0.28%)	0 (0.00%)	20 (0.00%)
U3423-1	40951542	40829476 (99.70%)	15596 (0.04%)	106454 (0.26%)	0 (0.00%)	16 (0.00%)
U3423-2	43854120	43714232 (99.68%)	23252 (0.05%)	116612 (0.27%)	0 (0.00%)	24 (0.00%)
U3423-3	39955312	39827988 (99.68%)	23998 (0.06%)	103326 (0.26%)	0 (0.00%)	0 (0.00%)
U6-1	40974348	40833328 (99.66%)	25276 (0.06%)	115730 (0.28%)	0 (0.00%)	14 (0.00%)

U6-2	46697160	46548458 (99.68%)	22422 (0.05%)	126256 (0.27%)	0 (0.00%)	24 (0.00%)
U6-3	54064266	53885742 (99.67%)	41442 (0.08%)	136928 (0.25%)	0 (0.00%)	154 (0.00%)
T15-1	48605880	48456232 (99.69%)	39572 (0.08%)	110056 (0.23%)	0 (0.00%)	20 (0.00%)
T15-2	49052402	48898474 (99.69%)	35946 (0.07%)	117960 (0.24%)	0 (0.00%)	22 (0.00%)
T15-3	42466136	42346634 (99.72%)	27200 (0.06%)	92264 (0.22%)	0 (0.00%)	38 (0.00%)

Note: Sample is the sample ID; RawDatas is the number of raw reads downloaded from the device; CleanData (%) is the number of high-quality reads and percentage (based on RawReads); Adapter (%) is the number of reads containing adapters and the percentage (based on RawReads); LowQuality (%) refers to low-quality reads. The number and percentage of reads with a quality value $Q \leq 20$ for more than 50% of the bases in single ended reads (based on RawReads); PolyA (%) is the number of polyA reads and the percentage (based on RawReads); N (%) is the number of high filtering reads with N ratio and the percentage (based on RawReads).

Table S3 Base Information Statistics Table

Sample	RawData(bp)	BF_Q20(%)	BF_Q30(%)	BF_N(%)	BF_GC(%)	CleanData(bp)	AF_Q20(%)	AF_Q30(%)	AF_N(%)	AF_GC(%)
H167-1	6259673100	6189581836 (98.88%)	6050856636 (96.66%)	60194 (0.00%)	3055811345 (48.82%)	6185077958	6127653797 (99.07%)	5998875952 (96.99%)	59226 (0.00%)	3015901019 (48.76%)
H167-2	6256078200	6153039286 (98.35%)	5975092353 (95.51%)	65297 (0.00%)	3064908972 (48.99%)	6161966425	6074518585 (98.58%)	5908080224 (95.88%)	63533 (0.00%)	3012122635 (48.88%)
H167-3	5619773100	5552542762 (98.80%)	5425154466 (96.54%)	123754 (0.00%)	2764272705 (49.19%)	5539957095	5486318795 (99.03%)	5369523766 (96.92%)	120312 (0.00%)	2720642458 (49.11%)
H74-1	10243581600	10097286928 (98.57%)	9821551154 (95.88%)	217967 (0.00%)	5025907051 (49.06%)	10098537704	9980122362 (98.83%)	9725393894 (96.30%)	211820 (0.00%)	4945980394 (48.98%)
H74-2	7210238100	7075893617 (98.14%)	6852030220 (95.03%)	126841 (0.00%)	3551467499 (49.26%)	7127759892	7009231706 (98.34%)	6795273701 (95.34%)	123114 (0.00%)	3505271837 (49.18%)
H74-3	10738435200	10616472974 (98.86%)	10377496330 (96.64%)	104009 (0.00%)	5290083207 (49.26%)	10600743444	10501869374 (99.07%)	10281441382 (96.99%)	102239 (0.00%)	5216618184 (49.21%)
U342-1	6142731300	6077287743 (98.93%)	5947827232 (96.83%)	60795 (0.00%)	3017205003 (49.12%)	6080516160	6026173016 (99.11%)	5905251034 (97.12%)	59897 (0.00%)	2983609459 (49.07%)
U342-2	6578118000	6506123438 (98.91%)	6363618920 (96.74%)	63500 (0.00%)	3238195302 (49.23%)	6506267046	6446938668 (99.09%)	6314505927 (97.05%)	62424 (0.00%)	3198585350 (49.16%)
U342-3	5993296800	5925251185 (98.86%)	5790699606 (96.62%)	28928 (0.00%)	2951668987 (49.25%)	5906445356	5852060538 (99.08%)	5728511221 (96.99%)	28463 (0.00%)	2904117281 (49.17%)
U6-1	6146152200	6074057077 (98.83%)	5935335968 (96.57%)	58520 (0.00%)	3048285451 (49.60%)	6056906306	5999976167 (99.06%)	5873075356 (96.96%)	57459 (0.00%)	2999408106 (49.52%)
U6-2	7004574000	6927416575 (98.90%)	6776844968 (96.75%)	68928 (0.00%)	3466225327 (49.49%)	6916462007	6854130204 (99.10%)	6715138463 (97.09%)	67698 (0.00%)	3418790869 (49.43%)

U6-3	8109639900	8012910696 (98.81%)	7829774866 (96.55%)	180432 (0.00%)	4024188190 (49.62%)	7980871613	7904592790 (99.04%)	7738213336 (96.96%)	175358 (0.00%)	3952569103 (49.53%)
T15-1	7290882000	7209573336 (98.88%)	7052843553 (96.74%)	71971 (0.00%)	3553187354 (48.73%)	7178515214	7115368740 (99.12%)	6973527095 (97.14%)	70462 (0.00%)	3492596029 (48.65%)
T15-2	7357860300	7277578241 (98.91%)	7119993735 (96.77%)	71863 (0.00%)	3579686349 (48.65%)	7263831866	7199415393 (99.11%)	7054867368 (97.12%)	70667 (0.00%)	3527741181 (48.57%)
T15-3	6369920400	6301435899 (98.92%)	6166379715 (96.80%)	102415 (0.00%)	3089957953 (48.51%)	6293877921	6237676937 (99.11%)	6112427139 (97.12%)	100736 (0.00%)	3048545361 (48.44%)

Note: Sample is the name of the sample; BF (Before filter) is the base information of the sample before filtering; AF (After filter) is the base information of the filtered sample; RawData (bp) is the total number of bases in the downloaded data (Unit bp); CleanData (bp) is the total number of bases in high-quality filtered data (Unit bp); Q20 (%) is the number of bases with sequencing base quality values above Q20 and the percentage of them in RawData (or CleanData); Q30 (%) is the number of bases with sequencing base quality values above Q30 and the percentage of them in RawData (or CleanData); N (%) is the number of N bases in a single ended read and the percentage of it in RawData (or CleanData); GC (%) is the proportion of GC bases in the sequence before (after) filtering.

Table S4 Alignment ribosome statistics

Sample	clean_reads	Mapped_Reads(%)	Unmapped_Reads(%)
H167-1	41597026	199528 (0.48%)	41397498 (99.52%)
H167-2	41500934	207520 (0.50%)	41293414 (99.50%)
H167-3	37334922	186710 (0.50%)	37148212 (99.50%)
H74-1	67974512	417820 (0.61%)	67556692 (99.39%)
H74-2	47833516	302714 (0.63%)	47530802 (99.37%)
H74-3	71355148	645324 (0.90%)	70709824 (99.10%)
U3423-1	40829476	238938 (0.59%)	40590538 (99.41%)
U3423-2	43714232	221606 (0.51%)	43492626 (99.49%)
U3423-3	39827988	254076 (0.64%)	39573912 (99.36%)
U6-1	40833328	362020 (0.89%)	40471308 (99.11%)
U6-2	46548458	378124 (0.81%)	46170334 (99.19%)
U6-3	53885742	1067772 (1.98%)	52817970 (98.02%)
T15-1	48456232	972886 (2.01%)	47483346 (97.99%)
T15-2	48898474	1031604 (2.11%)	47866870 (97.89%)
T15-3	42346634	318322 (0.75%)	42028312 (99.25%)

Note: Sample is the sample name; Clean_reads is the number of high-quality reads; Mapped_Reads (%) is the number and percentage of reads (based on clean reads) compared to the ribosomes of this species; Unmapped_Reads (%) is the number of reads that cannot be aligned with ribosomes and the percentage based on clean reads.

Table S5 Alignment of reference statistics

Sample	Total	Unmapped(%)	Unique_Mapped (%)	Multiple_Mapped (%)	Total_Mapped(%)
H167-1	41397498	6638723 (16.04%)	33522923 (80.98%)	1235852 (2.99%)	34758775 (83.96%)
H167-2	41293414	6508677 (15.76%)	33618896 (81.41%)	1165841 (2.82%)	34784737 (84.24%)
H167-3	37148212	5531447 (14.89%)	30492378 (82.08%)	1124387 (3.03%)	31616765 (85.11%)
H74-1	67556692	10058113 (14.89%)	55428636 (82.05%)	2069943 (3.06%)	57498579 (85.11%)
H74-2	47530802	6778032 (14.26%)	39313709 (82.71%)	1439061 (3.03%)	40752770 (85.74%)
H74-3	70709824	10113427 (14.30%)	58305820 (82.46%)	2290577 (3.24%)	60596397 (85.70%)
U3423-1	40590538	5647225 (13.91%)	33753890 (83.16%)	1189423 (2.93%)	34943313 (86.09%)
U3423-2	43492626	5938218 (13.65%)	36266688 (83.39%)	1287720 (2.96%)	37554408 (86.35%)

U3423-3	39573912	5398141 (13.64%)	32990426 (83.36%)	1185345 (3.00%)	34175771 (86.36%)
U6-1	40471308	5946955 (14.69%)	33296587 (82.27%)	1227766 (3.03%)	34524353 (85.31%)
U6-2	46170334	6795909 (14.72%)	37957579 (82.21%)	1416846 (3.07%)	39374425 (85.28%)
U6-3	52817970	7767967 (14.71%)	43130980 (81.66%)	1919023 (3.63%)	45050003 (85.29%)
T15-1	47483346	7816610 (16.46%)	37933323 (79.89%)	1733413 (3.65%)	39666736 (83.54%)
T15-2	47866870	7912497 (16.53%)	38165331 (79.73%)	1789042 (3.74%)	39954373 (83.47%)
T15-3	42028312	6977344 (16.60%)	33753784 (80.31%)	1297184 (3.09%)	35050968 (83.40%)

Note: Sample is the sample name; Total is the number of reads after filtering ribosomes, known as effective reads; Unmapped (%) refers to the number of reads that have not been aligned with the reference genome and the proportion of effective reads; Unique_Mapped (%) is the number of reads and the proportion of valid reads that are uniquely aligned with the reference genome; Multiple_Mapped (%) refers to the number of reads and the proportion of effective reads on the reference genome compared at multiple locations; Total_Mapped (%) is the total number of reads that can be located on the genome and the proportion of effective reads.

Table S6 Align the statistical situation of the reference area

sample	exon	intron	intergenic
H167-1	31032085 (89.28%)	2975531 (8.56%)	751159 (2.16%)
H167-2	31173635 (89.62%)	2879186 (8.28%)	731916 (2.10%)
H167-3	28436201 (89.94%)	2520824 (7.97%)	659740 (2.09%)
H74-1	50962373 (88.63%)	5254562 (9.14%)	1281644 (2.23%)
H74-2	36442928 (89.42%)	3431200 (8.42%)	878642 (2.16%)
H74-3	54206766 (89.46%)	5003492 (8.26%)	1386139 (2.29%)
U3423-1	31305294 (89.59%)	2872505 (8.22%)	765514 (2.19%)
U3423-2	33691031 (89.71%)	3055756 (8.14%)	807621 (2.15%)
U3423-3	30622322 (89.60%)	2800963 (8.20%)	752486 (2.20%)
U6-1	30949705 (89.65%)	2827539 (8.19%)	747109 (2.16%)
U6-2	35300008 (89.65%)	3226347 (8.19%)	848070 (2.15%)
U6-3	39690801 (88.10%)	4151243 (9.21%)	1207959 (2.68%)
T15-1	34731960 (87.56%)	3760363 (9.48%)	1174413 (2.96%)
T15-2	34838736 (87.20%)	3890991 (9.74%)	1224646 (3.07%)
T15-3	31249454 (89.15%)	2972689 (8.48%)	828825 (2.36%)

Table S7 New gene annotation table

GeneID	Symbol	Description	KEGG_A_class	KEGG_B_class	Pathway	K_ID	GO Component	GO Function	GO Process	TF_family
MSTRG.100	-	KAK3444190.1 taxid=71139 hypothetical protein EUGRSUZ_A00032 [Eucalyptus grandis]	-	-	-	-	-	GO:0003677//DNA binding	GO:0006355//regulation of transcription, DNA-templated	-
MSTRG.1	-	-	-	-	-	-	-	GO:0033926//glycopeptide alpha-N-acetylgalactosaminidase activity	-	-

MSTRG.10001010	CODM	XP_018727310.2 taxid=71139 probable 2-oxoglutarate-dependent dioxygenase ANS [Eucalyptus grandis]	-	-	-	-	GO:0005886//plasma membrane	-	-	-
MSTRG.10001012	--	KAK3439526.1 taxid=71139 hypothetical protein EUGRSUZ_C04396 [Eucalyptus grandis]	-	-	-	-	-	-	-	-
MST	--	KAK3439535.1 taxid=71139 hypothetical	-	-	-	-	-	-	-	-

RG.10026	-	protein EUGRSUZ_C04401 [Eucalyptus grandis]	-	-	-	-	-					
MSTRG.10032	-	XP_039168969.1 taxid=71139 receptor-like protein 15 [Eucalyptus grandis]	-	-	-	-	-		GO:0005515//protein binding	-		-
MSTRG.10044	-	XP_039167687.1 taxid=71139 28S ribosomal protein S33, mitochondrial [Eucalyptus grandis]	-	-	-	-	GO:0000313//organellar ribosome;GO:0000314//organellar small ribosomal subunit;GO:0005575//cellular_component;GO:0005622//intracellular;GO:0005623//cell;GO:0005737//cytoplasm;GO:0005739//mitochondrion;GO:0005759//mitochondrial matrix;GO:0005761//mitochondrial ribosome;GO:0005763//mitochondrial small	GO:0003674//molecular_function;GO:0003735//structural constituent of ribosome;GO:0005198//structural molecule activity	GO:0006412//translation;GO:0006518//peptide metabolic process;GO:0006807//nitrogen compound metabolic process;GO:0008150//biological_process;GO:0008152//metabolic process;GO:0009058//biosynthetic process;GO:0009059//macromolecule biosynthetic process;GO:0009987//cellular process;GO:0010467//gene	-	-

1	ribosomal subunit;GO:0005840//ribosome;GO:0015935//small ribosomal subunit;GO:0031974//membrane-enclosed lumen;GO:0032991//macromolecular complex;GO:0043226//organelle;GO:0043227//membrane-bounded organelle;GO:0043228//non-membrane-bounded organelle;GO:0043229//intracellular organelle;GO:0043231//intracellular membrane-bounded organelle;GO:0043232//intracellular non-membrane-bounded organelle;GO:0043233//organelle lumen;GO:0044391//ribosomal subunit;GO:0044422//organelle part;GO:0044424//intracellular part;GO:0044429//mitochondrial part;GO:0044444//cytoplasmic part;GO:0044446//intracellular organelle part;GO:0044464//cell part;GO:0070013//intracellular organelle lumen;GO:0098798//mitochondrial protein complex;GO:1990904//ribonucleoprotein complex	expression;GO:0019538//protein metabolic process;GO:0032543//mitochondrial translation;GO:0034641//cellular nitrogen compound metabolic process;GO:0034645//cellular macromolecule biosynthetic process;GO:0043043//peptide biosynthetic process;GO:0043170//macromolecule metabolic process;GO:0043603//cellular amide metabolic process;GO:0043604//amide biosynthetic process;GO:0044237//cellular metabolic process;GO:0044238//primary metabolic process;GO:0044249//cellular biosynthetic process;GO:0044260//cellular macromolecule metabolic process;GO:0044267//cellular protein metabolic process;GO:0044271//cellular nitrogen compound biosynthetic process;GO:0071704//organic substance metabolic process;GO:1901564//organonitrogen compound metabolic process;GO:1901566//organonitrogen compound biosynthetic process;GO:1901576//organic substance biosynthetic process

MSTRG.10006	-	XP_039163050.1 taxid=71139 LOW QUALITY PROTEIN: G-type lectin S-receptor-like serine/threonine-protein kinase LECRK3 [Eucalyptus grandis]	-	-	-	-	-		-	GO:0050830//defense response to Gram-positive bacterium	-
MSTRG.10076	-	-	-	-	-	-	-		-	-	-
MSTRG	-	-	-	-	-	-	GO:0015935//small ribosomal subunit	GO:0003735//structural constituent of ribosome		GO:0006412//translation	-

.1
0
0
8
9

Note: GeneID is a new gene ID; Symbol is the name of a gene; Description is an Nr annotation; KEGG_A_class is the first level annotation of KEGG; KEGG_B_class is the second level annotation of KEGG; Pathway is a KEGG pathway annotation; GO Component is a component annotation for GO cells; GO Function is a functional annotation of GO molecules; GO Process is a GO biological process annotation; TF_family is a transcription factor annotation (This table only displays the top ten new genes. Please refer to the complete Excel spreadsheet 'New Gene Annotation' for all new genes.).

Table S8 Genetic testing statistics

sample	Refer_Genes	sequenced_Refer_Genes(%)	Novel_Genes	sequenced_Novel_Genes(%)	Total_Genes	sequenced_Total_Genes(%)
all	36005	30215 (83.92%)	3071	3071 (100.00%)	39076	33286 (85.18%)
H167-1	36005	24903 (69.17%)	3071	2260 (73.59%)	39076	27163 (69.51%)
H167-2	36005	24404 (67.78%)	3071	2228 (72.55%)	39076	26632 (68.15%)
H167-3	36005	24662 (68.50%)	3071	2233 (72.71%)	39076	26895 (68.83%)
H74-1	36005	24577 (68.26%)	3071	2124 (69.16%)	39076	26701 (68.33%)
H74-2	36005	25539 (70.93%)	3071	2292 (74.63%)	39076	27831 (71.22%)
H74-3	36005	24079 (66.88%)	3071	2163 (70.43%)	39076	26242 (67.16%)
U3423-1	36005	23151 (64.30%)	3071	2123 (69.13%)	39076	25274 (64.68%)
U3423-2	36005	23213 (64.47%)	3071	2095 (68.22%)	39076	25308 (64.77%)
U3423-3	36005	23112 (64.19%)	3071	2101 (68.41%)	39076	25213 (64.52%)
U6-1	36005	24650 (68.46%)	3071	2132 (69.42%)	39076	26782 (68.54%)
U6-2	36005	24724 (68.67%)	3071	2148 (69.94%)	39076	26872 (68.77%)
U6-3	36005	25033 (69.53%)	3071	2183 (71.08%)	39076	27216 (69.65%)
T15-1	36005	24301 (67.49%)	3071	2166 (70.53%)	39076	26467 (67.73%)
T15-2	36005	24098 (66.93%)	3071	2129 (69.33%)	39076	26227 (67.12%)
T15-3	36005	23943 (66.50%)	3071	2089 (68.02%)	39076	26032 (66.62%)

Note: Refer_Genes is the total number of genes in the reference genome (or set of reference genes), and when there are no new genes, there are no results in this column; Sequenced_Refer_Genes (%) is the total number and percentage of Referr_Genes detected in the

sequencing results. When there are no new genes present, there is no result in this column; Novel_Genes is the number of new genes discovered in the project. When there are no new genes, there will be no results in this column; Sequenced_Novel_Genes(%) is the total number and percentage of Novel Genes detected in the sequencing results. When there are no new genes present, there will be no results in this column; Total_Genes is the total number of genes, including the reference genome and new genes; Sequenced_Total_Genes(%) is the total number of genes detected in the sequencing results and the percentage of them to All_Genes.

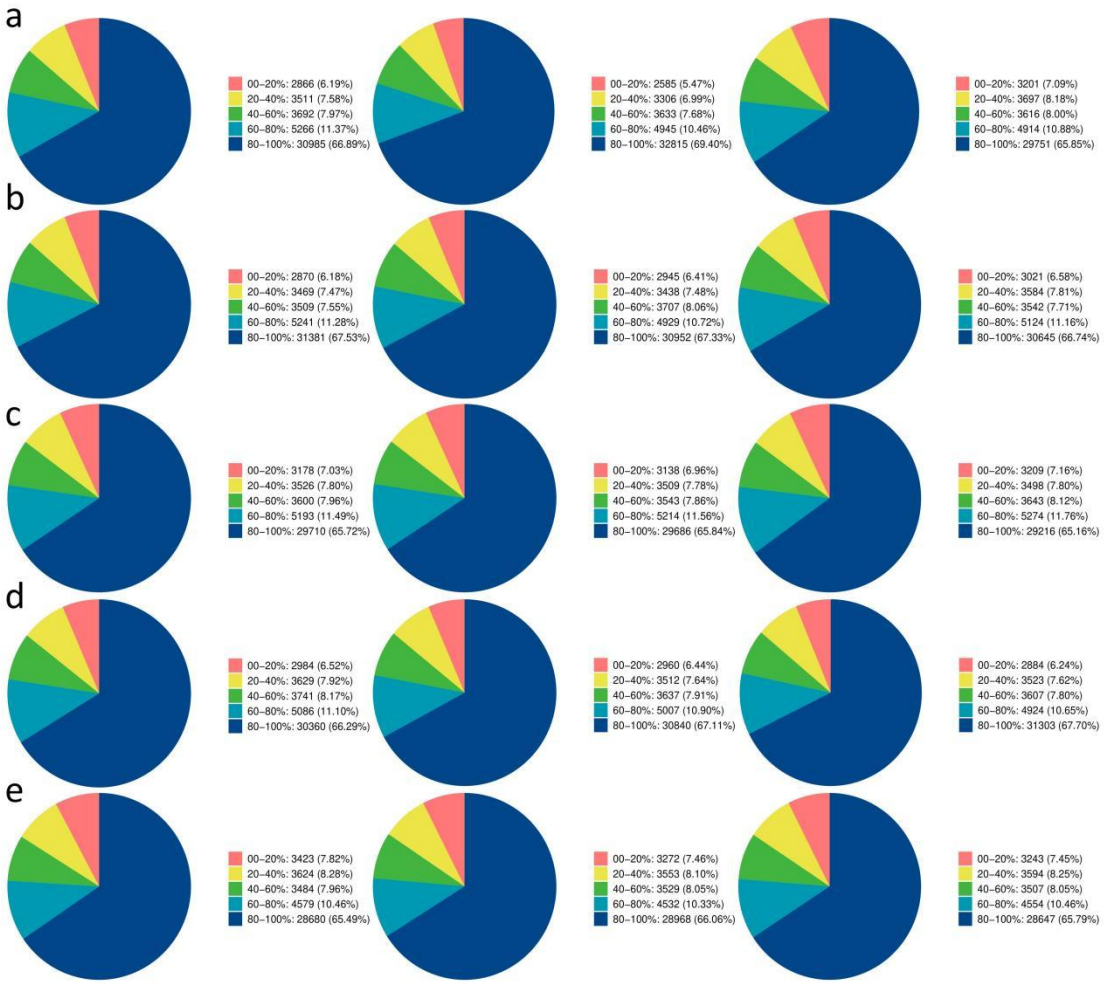


Figure S4 Gene coverage distribution (Note: The different colors in the figure represent the proportion of genes with a certain coverage range. a: H74; b: H167; c: T15; d: U6; e: U3423.)

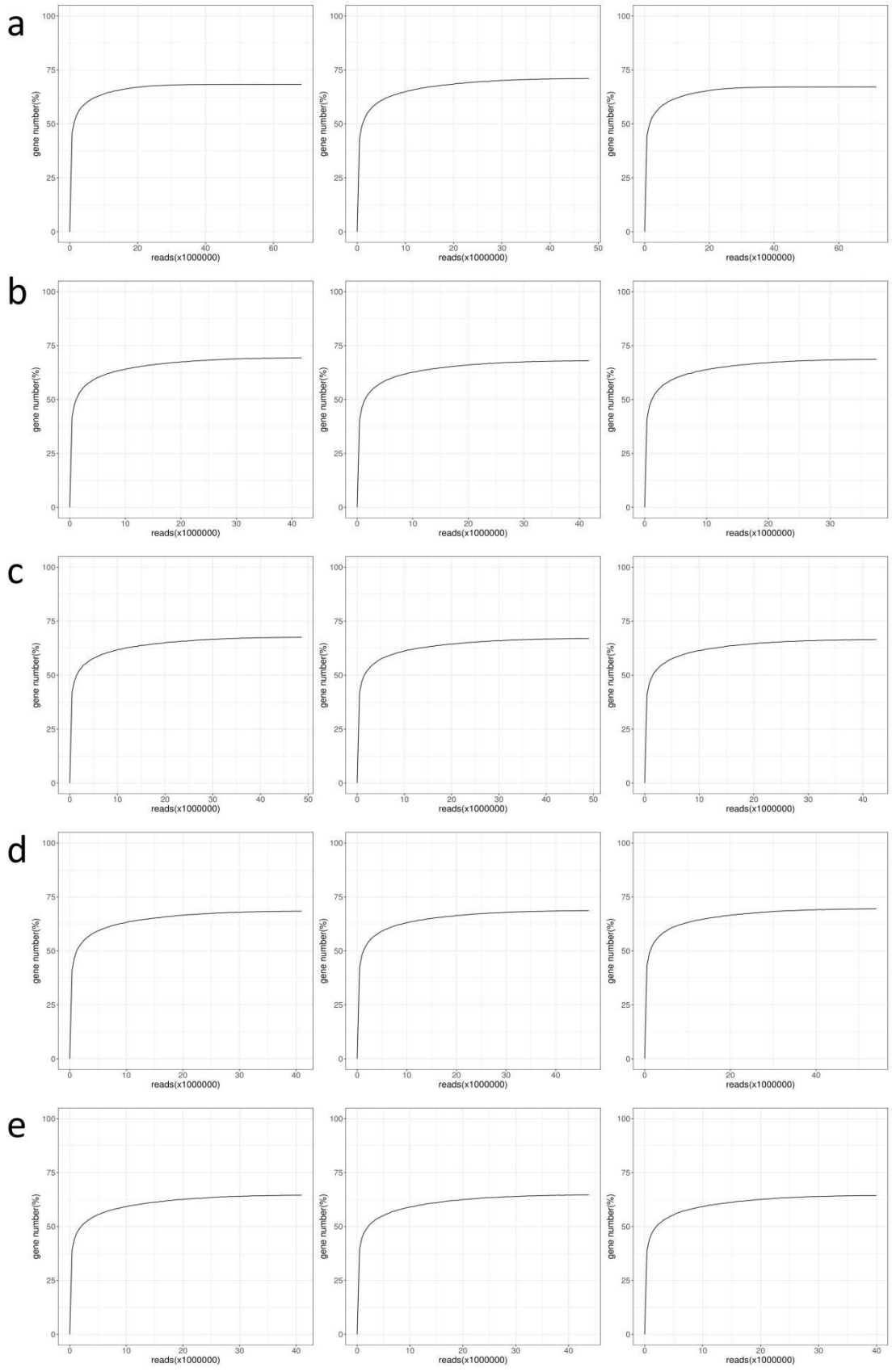


Figure S5 Sequencing saturation distribution (Note: a: H74; b: H167; c: T15; d: U6; e: U3423.)

Table S9 Gene expression level table

Id	H167-1_tpm	H167-2_tpm	H167-3_tpm	H774-1_tpm	H774-2_tpm	H774-3_tpm	U3423-1_tpm	U3423-2_tpm	U3423-3_tpm	U66-1_tpm	U66-2_tpm	U66-3_tpm	T115-1_tpm	T115-2_tpm	T115-3_tpm	H167_count	H167-1_count	H167-2_count	H167-3_count	H774_count	H774-1_count	H774-2_count	H774-3_count	U3423_count	U3423-1_count	U3423-2_count	U3423-3_count	U66_count	U66-1_count	U66-2_count	U66-3_count	T115_count	T115-1_count	T115-2_count	T115-3_count	Symbol	Description	KEGG_Class	KEGG_Blast	Pathway	K_ID	GO Component	GO Function	GO Process	TF_family
MSTRG.100	0	0	0	0	0	0	0	0	0	0	0	0	0.397	0.274	0.271	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	1	-	KAK34441 90.1 taxid=71139 hypothetical protein EUG	-	-	-	-	-	GO:0003677//DNA binding	GO:0006355//regulation of transcription, DNA-templated	-

																																RSUZ_A00032 [Eucalyptus grandis]								
MSTRG.10000	3.001	0.51	3.102	11.347	19.871	7.552	0.49	0	0.244	1.037	0.173	1.055	7.358	8.114	6.872	35.45	5.97	34.9	214.88	282.53	163.43	6.18	0	3	13.29	2.5	17.42	101.32	110.15	82.97	--	-	-	-	-	-	-	GO:0033926//glycopeptide alpha-N-acetylgalactosaminidase activity	-	--
MSTR	0	0.29	0	0	0	0	0	0	0	0	0	0	0.26	0	0.57	0	1	0	0	0	0	0	0	0	0	0	0	1.03	0	2	CODM	XP_018727310.2 taxid=	-	-	-	-	GO:0005886//plasma membrane	-	-	--

G.100010		7													5																71139 probable 2-oxoglutarate-dependent dioxygenase ANS [Eucalyptus grandis]								
MSTRG.1001	8.844	4.782	5.909	11.366	12.987	11.604	12.303	9.716	12.698	5.515	7.574	7.796	6.054	6.14	6.117	99	53	63	204	175	238	147	124	148	67	104	122	79	79	70	--	KAK3439526.1 taxid=71139 hypothetical protein	-	-	-	-	-	-	-	--

2																																											
MSTRG.10026	0	0	0	0	0	0	0	0	0	0	0	0	1.048	0.354	0.797	0	0	0	0	0	0	0	0	0	0	0	0	3	1	2	--	EUGRSUZ_C04396 [Eucalyptus grandis] KAK3439535.1 taxid=71139 hypothetical protein EUGRSUZ_C04401 [Eucalyptus grandis]	-	-	-	-	-		-		-		--

MSTRG.100032	0	0	0	0	1.138	0.299	0	0	0	6.126	5.135	3.333	0	0	0	0	0	0	0	5	2	0	0	0	24.28	23	17	0	0	0	--	XP_039168969.1 taxid=71139 receptor-like protein 15 [Eucalyptus grandis]	-	-	-	-	-	GO:0005515//protein binding	-	--
MSTRG.100041	25.578	25.834	35.234	18.124	28.73	24.348	35.083	22.794	29.677	21.306	24.522	25.469	50.378	35.252	45.042	125	125	164	142	169	218	183	237	151	113	147	174	287	198	225	--	XP_039167687.1 taxid=71139 28S ribosomal protein S33, mitochondr	-	-	-	-	GO:0000313//organellar ribosome;GO:0000314//organellar small ribosomal subunit;GO:0005575//cellular_component;GO:0005622//intracellular;GO:0005623//cell;G	GO:0003674//molecular_function;GO:0003735//structural	GO:0006412//translation;GO:0006518//peptide metabolic process;GO:0006807//nitrogen compound metabolic process;GO:0008150//biological_process;GO:0008152//meta	--

ial [Eucalyptus grandis]	O:0005737//cytoplasm;GO:0005739//mitochondrion;GO:0005759//mitochondrial matrix;GO:0005761//mitochondrial ribosome;GO:0005763//mitochondrial small ribosomal subunit;GO:0005840//ribosome;GO:0015935//small ribosomal subunit;GO:0031974//membrane-enclosed lumen;GO:0032991//macromolecular complex;GO:0043226//organelle	constituent of ribosome;GO:0005198//structural molecule activity	bolic process;GO:0009058//biosynthetic process;GO:0009059//macromolecule biosynthetic process;GO:0009987//cellular process;GO:0010467//gene expression;GO:0019538//protein metabolic process;GO:0032543//mitochondrial translation;GO:0034641//cellular nitrogen compound metabolic process;GO:0034645//cellular macromolecule

lle;GO:0043227//membrane-bounded organelle;GO:0043228//non-membrane-bounded organelle;GO:0043229//intracellular organelle;GO:0043231//intracellular membrane-bounded organelle;GO:0043232//intracellular non-membrane-bounded organelle;GO:0043233//organelle lumen;GO:0044391//ribosomal

biosynthetic process;GO:0043043//peptide biosynthetic process;GO:0043170//macromolecule metabolic process;GO:0043603//cellular amide metabolic process;GO:0043604//amide biosynthetic process;GO:0044237//cellular metabolic process;GO:0044238//primary metabolic process;GO:0044249//cellular biosynthetic process;GO:0044260//cellular

subunit;GO:00
44422//organell
e
part;GO:00444
24//intracellula
r
part;GO:00444
29//mitochondr
ial
part;GO:00444
44//cytoplasmi
c
part;GO:00444
46//intracellula
r organelle
part;GO:00444
64//cell
part;GO:00700
13//intracellula
r organelle
lumen;GO:009
8798//mitochon
drial protein
complex;GO:1
990904//ribonu
cleoprotein

macromolecule
metabolic
process;GO:00
44267//cellular
protein
metabolic
process;GO:00
44271//cellular
nitrogen
compound
biosynthetic
process;GO:00
71704//organic
substance
metabolic
process;GO:19
01564//organon
itrogen
compound
metabolic
process;GO:19
01566//organon
itrogen
compound
biosynthetic
process;GO:19

																																				complex		01576//organic substance biosynthetic process	
MSTRG.100068	0	0	0	0	0	0	0	0	0	0	0	0	0.699	0.236	1.063	0	0	0	0	0	0	0	0	0	0	0	0	3	1	4	--	XP_039163050.1 taxid=71139 LOW QUALITY PROTEIN: G-type lectin S-receptor-like serine/threonine-protein kinase LECR	-	-	-	-	-	-	GO:0050830//defense response to Gram-positive bacterium	--

																																K3 [Eucalyptus grandis]								
MSTRG.10076	22.269	22.646	22.844	17.792	18.296	46.644	65.863	73.022	68.425	13.372	11.263	14.241	20.716	21.336	21.071	75.82	76.34	74.08	97.12	74.98	290.96	239.35	283.45	242.55	49.41	47.04	67.78	82.22	83.49	73.33	--	-	-	-	-	-	-	-	-	--
MSTRG.10008	0	0	0	5.182	2.655	1.628	0	0	0	0	0	0	0	0	0	0	0	0	78	30	28	0	0	0	0	0	0	0	0	0	--	-	-	-	-	-	GO:0015935//small ribosomal subunit	GO:0003735//structural constituent of ribosome	GO:0006412//translation	--

Note: Id is the gene ID; Xx_tpm is the tpm expression level of the corresponding gene for each sample; Xx_count is the original reads count of the corresponding genes for each sample; Symbol is the name of a gene; Description is an Nr annotation; KEGG_A_class is the first level annotation of KEGG; KEGG_B_class is the second level annotation of KEGG; Pathway is a KEGG pathway annotation; GO Component is a component annotation for GO cells; GO Function is a functional annotation of GO molecules; GO Process is a GO biological process annotation; TF_family is a transcription factor annotation (This table only displays the top ten genes. Please refer to the complete Excel spreadsheet 'Gene expression level table' for all new genes.).

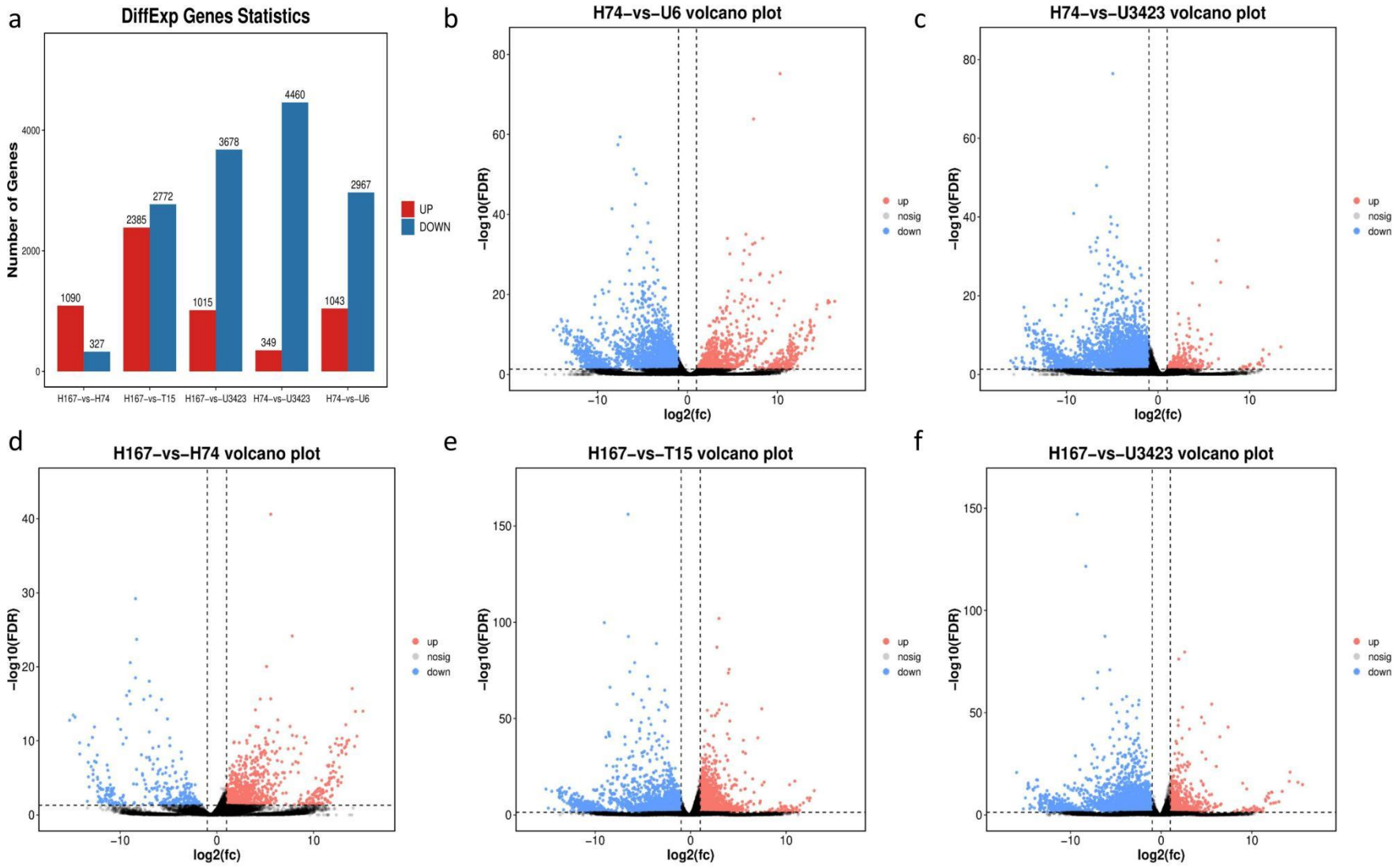

a
DiffExp Genes Statistics
Number of Genes
4000
2000
0
1090
327
2385
2772
1015
3678
349
4460
1043
2967
UP
DOWN
H167−vs−H74
H167−vs−T15
H167−vs−U3423
H74−vs−U3423
H74−vs−U6
b
H74−vs−U6 volcano plot
−log10(FDR)
log2(fc)
up
nosig
down
c
H74−vs−U3423 volcano plot
d
H167−vs−H74 volcano plot
e
H167−vs−T15 volcano plot
f
H167−vs−U3423 volcano plot

Figure S6 Analysis of inter group differences (Note: a: Differential gene statistical chart (X-axis: Samples compared in pairs; Y-axis: number of differentially expressed genes; Red represents upregulated differentially expressed genes; Blue represents differentially expressed genes that are downregulated.); b-f: Comparison of volcano maps with differences (The X-axis represents the logarithmic value of the difference multiple between two groups, and the Y-axis represents the negative Log10 value of the FDR of the difference between the two groups. The red (up-regulated expression level of group_2 relative to group_1) and blue (down regulated expression level) dots indicate differences in gene expression levels (with FDR<0.05 and a difference factor of more than twice), while the black dots indicate no difference.).)

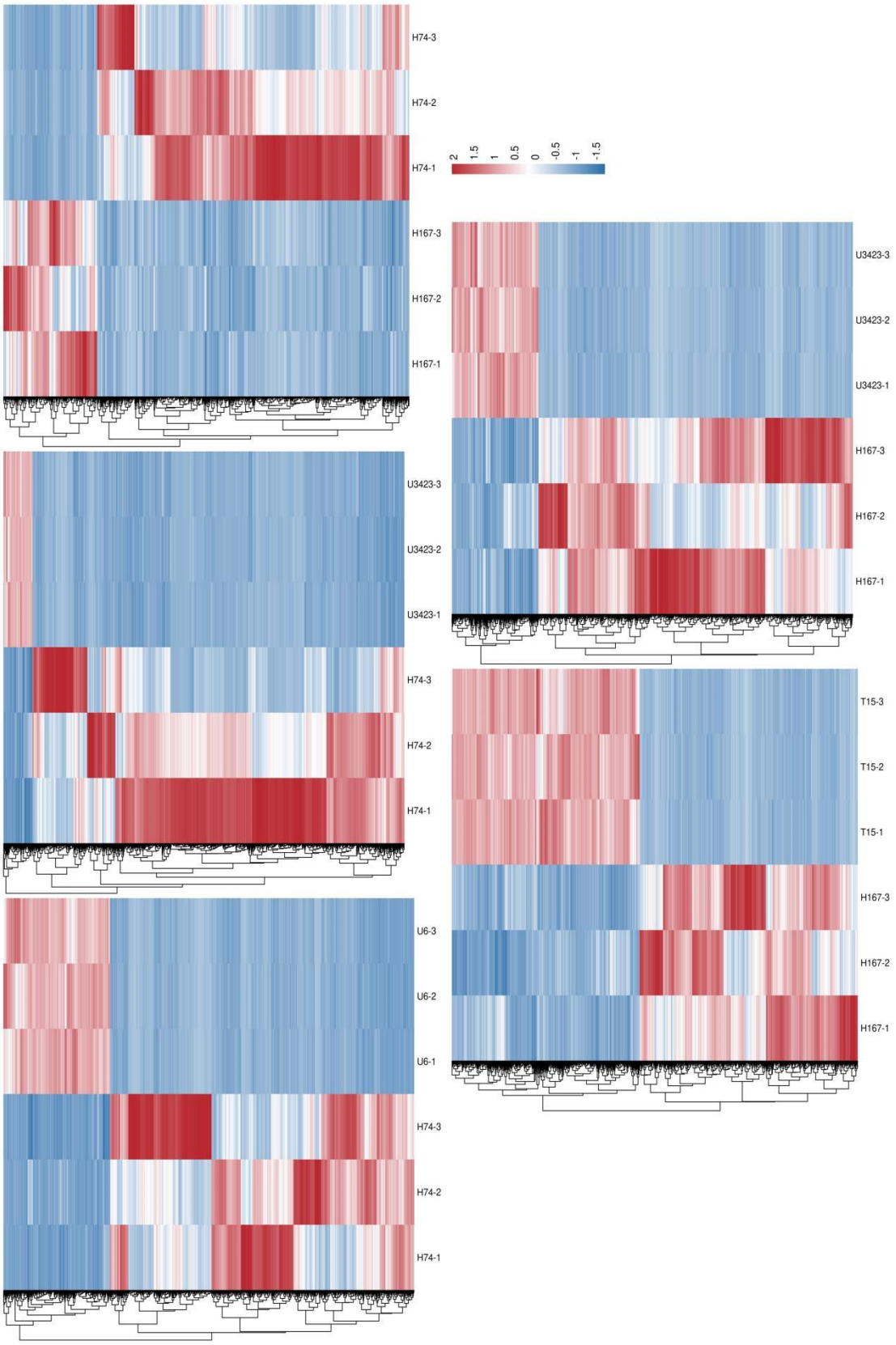

H74-3
H74-2
H74-1
H167-3
H167-2
H167-1
2
1.5
1
0.5
0
-0.5
-1
-1.5
U3423-3
U3423-2
U3423-1
H74-3
H74-2
H74-1
U3423-3
U3423-2
U3423-1
H167-3
H167-2
H167-1
U6-3
U6-2
U6-1
H74-3
H74-2
H74-1
T15-3
T15-2
T15-1
H167-3
H167-2
H167-1

Figure S7 Differential comparison clustering heatmap (Note: Each column in the figure represents a sample, and each row represents a gene. The expression levels of genes in different samples are represented by different colors, with red indicating higher expression levels and blue indicating lower expression levels.)

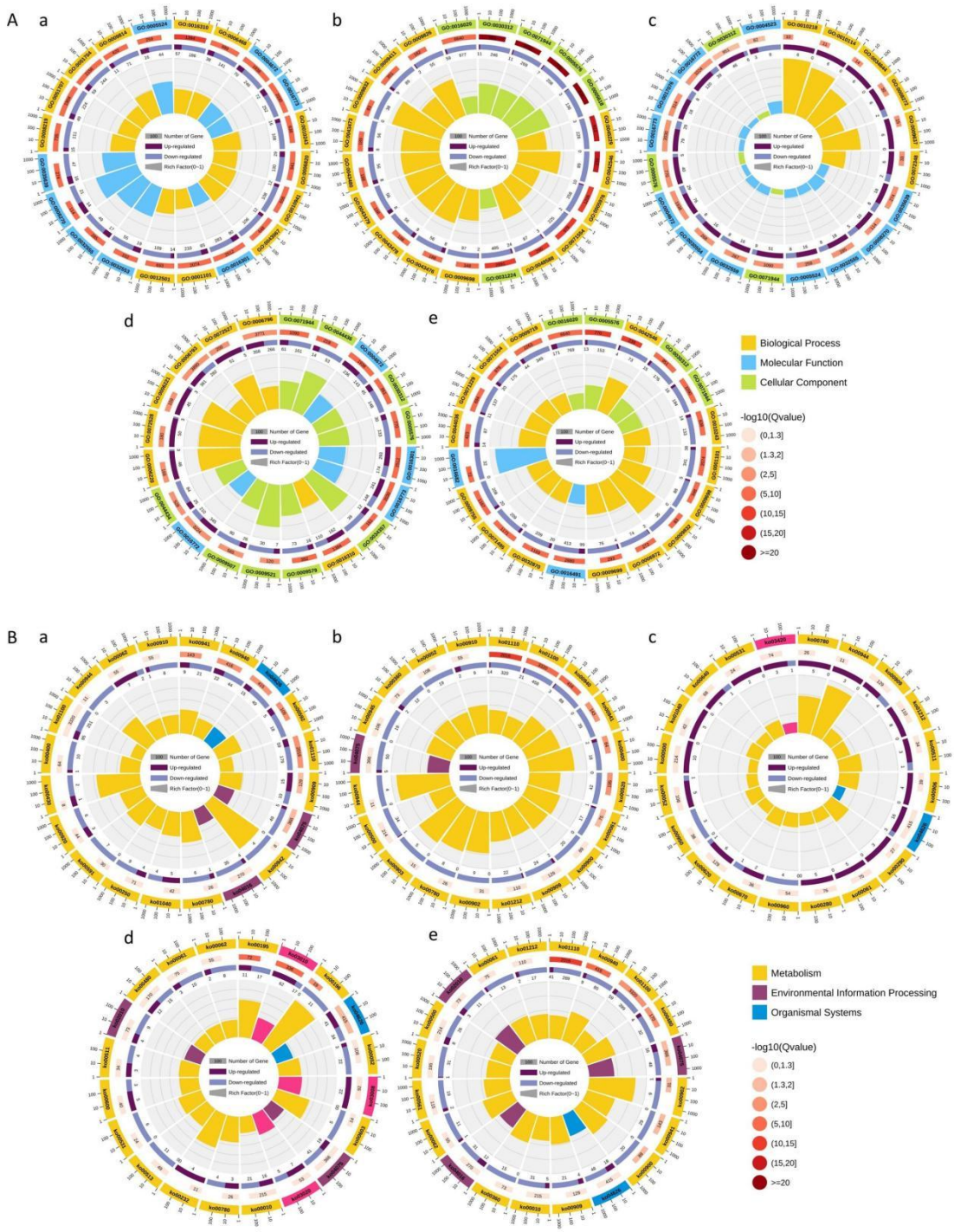


Figure S8 Enrichment circle diagram (Note: A: GO; B: KEGG. a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs T15; e: H167 vs U3423. First circle: Enrich the top 20 GOterms (A)/pathway (B), and outside the circle is the coordinate ruler of the number of differentially expressed genes. Different colors represent different Ontology (A)/A class (B); Second circle: The number and Q value of the GOterm (A)/pathway (B) in the differential gene background. The more differential gene backgrounds there are, the longer the bars, and the smaller the Q value, the redder the color; Third circle: Bar chart of the proportion of upregulated and downregulated differentially expressed genes, with dark purple representing the proportion of upregulated differentially expressed genes and light purple representing the

proportion of downregulated differentially expressed genes; The specific numerical values are displayed below; Fourth circle: RichFactor values for each GOterm (A)/pathway (B) (the number of differentially expressed genes in the GOterm (A)/pathway (B) divided by all numbers in the GOterm (A)/pathway (B)), background grid lines, with each grid representing 0.1.)

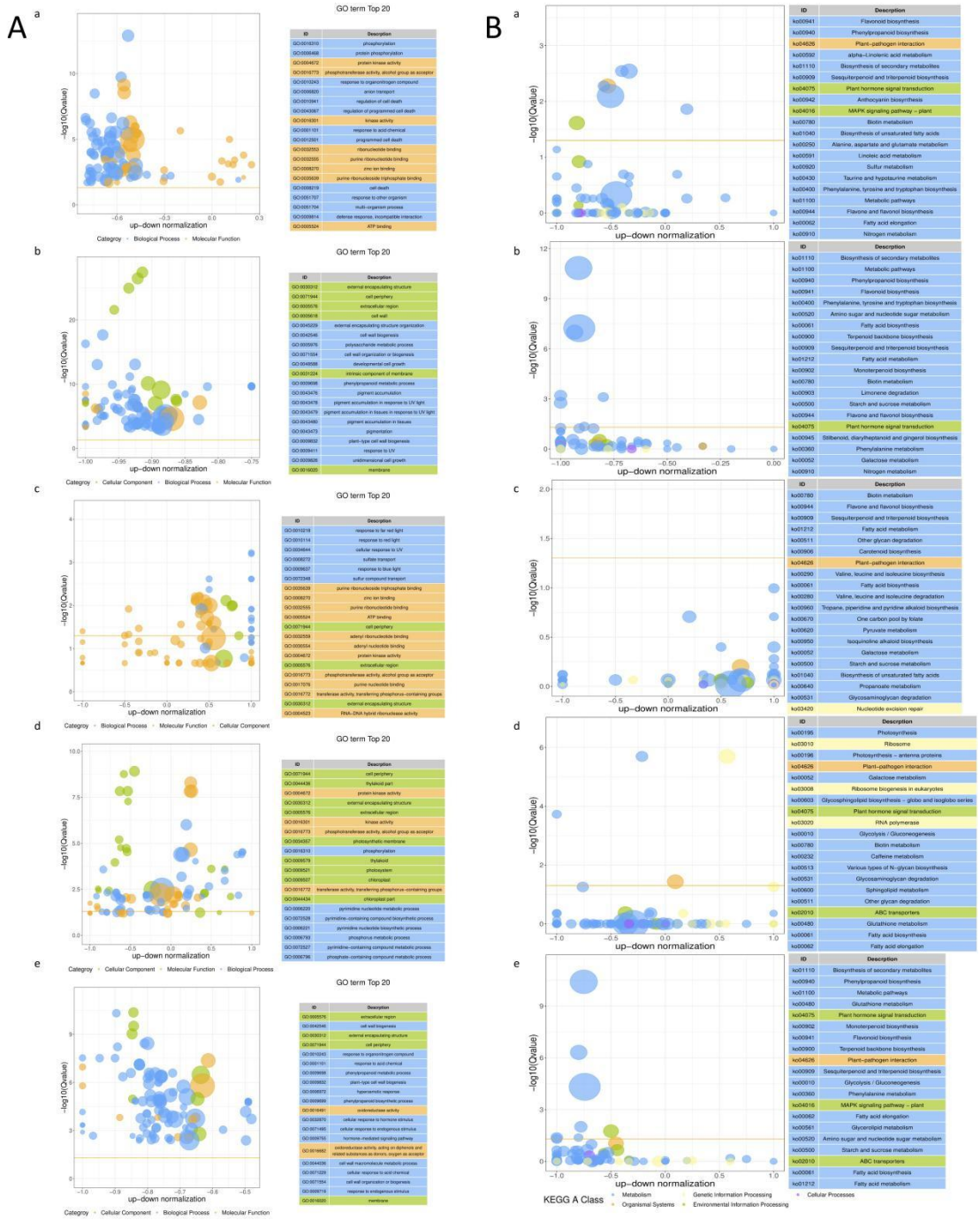


Figure S9 Enrichment difference bubble diagram (Note: A: GO; B: KEGG. a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs T15; e: H167 vs U3423. The vertical axis is -log10 (Qvalue), and the horizontal axis is the z-score value (the ratio of the difference between the number of up-regulated and down regulated differentially expressed genes to the

total differentially expressed genes). The yellow line represents the threshold of Qvalue=0.05. On the right is the top 20 GOterm (A)/pathway (B) list of Q values. Different colors represent different Ontology (A)/A class (B); Draw separate graphs for different Ontology (A)/A class (B).)

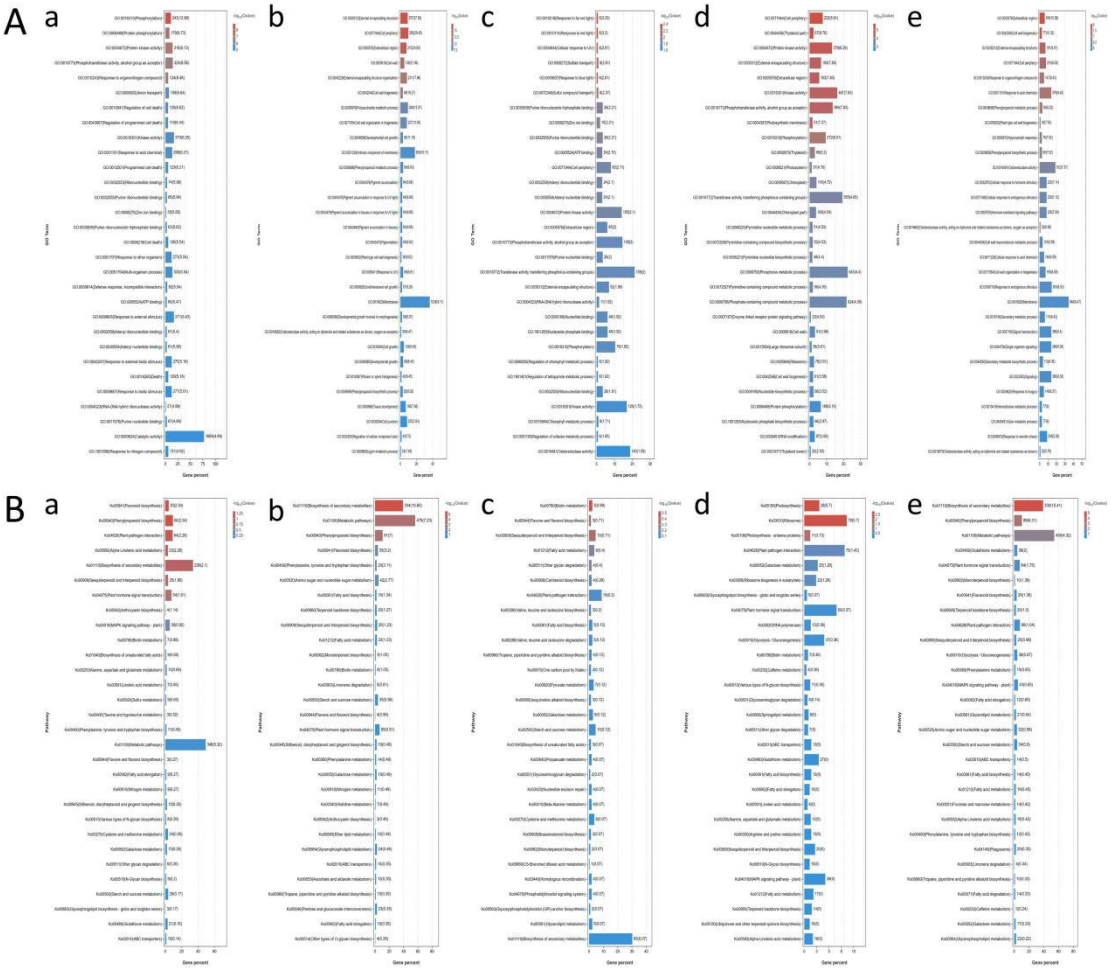


Figure S10 Enrichment bar diagram (Note: A: GO; B: KEGG. a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs T15; e: H167 vs U3423. Use the top 20 GOterms (A)/pathway (B) with the lowest Q value to plot, with the vertical axis representing the number of GOterms (A)/pathway (B) and the horizontal axis representing the percentage of this GOterm (A)/pathway (B) to the total number of differentially expressed genes. The darker the color, the smaller the Q value. The values on the column represent the number of GOterms (A)/pathway (B) and their Q values.)

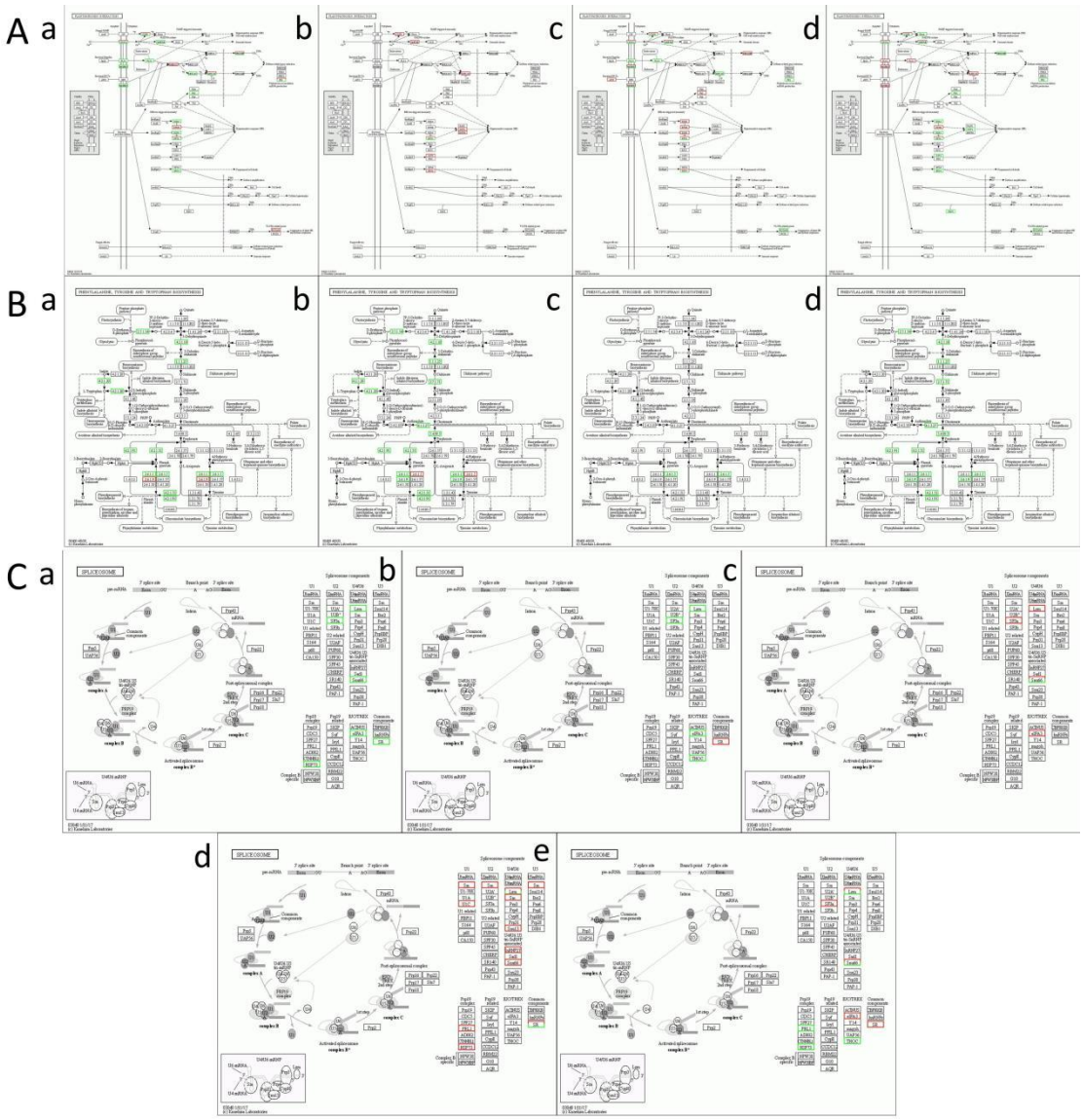


Figure S11 Detailed information on Pathways in KEGG database (Note: A: Plant-pathogen interaction (KO04626); a: H74 vs U6; b: H167 vs H74; c: H167 vs T15; d: H167 vs U3423; B: Phenylalanine, tyrosine and tryptophan biosynthesis (KO00400); a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs U3423; C: Spliceosome (KO03040); a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs T15; e: H167 vs U3423. The location of upregulated gene products is marked in red, while the location of downregulated gene products is marked in green.)

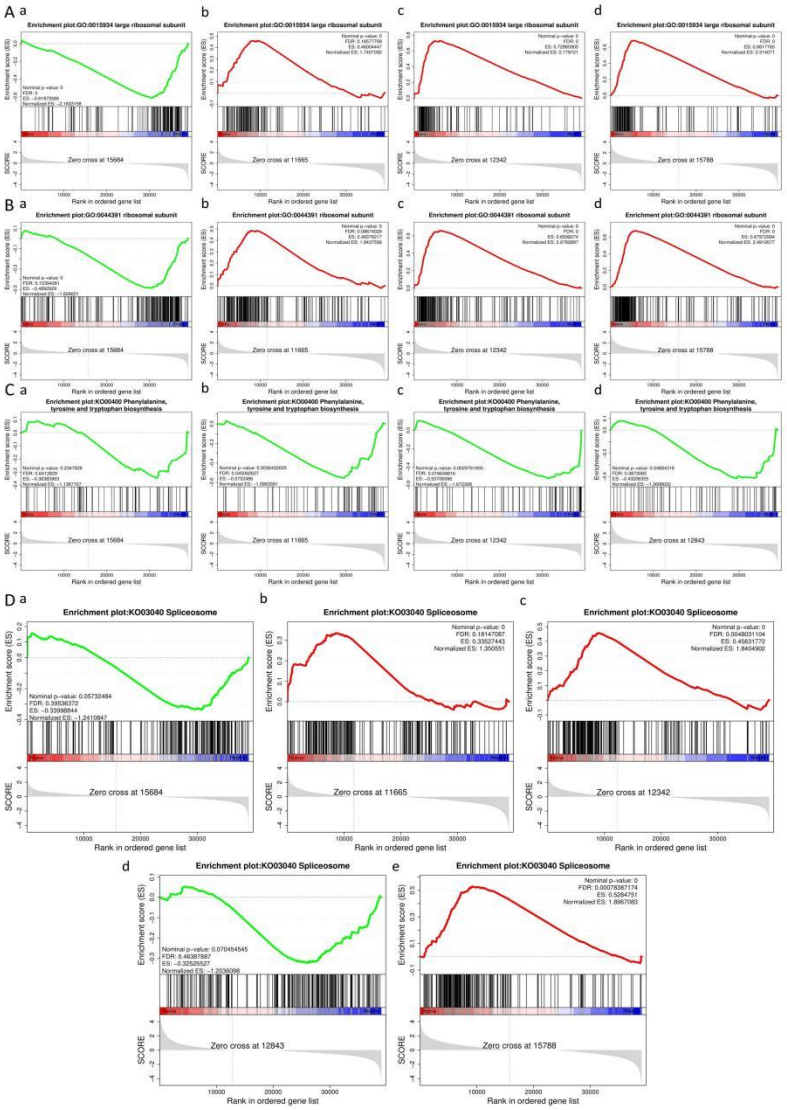


Figure S12 GSEA ES diagram (Note: A: GO:0015934; a: H167-H74, b: H167-U3423, c: H74-U3423, d: H167-T15. B: GO:0044391; a: H167-H74, b: H167-U3423, c: H74-U3423, d: H167-T15. C: KO00400; a: H167-H74, b: H167-U3423, c: H74-U3423, d: H74-U6. D: KO03040; a: H167-H74, b: H167-U3423, c: H74-U3423, d: H74-U6, e: H167-T15.)

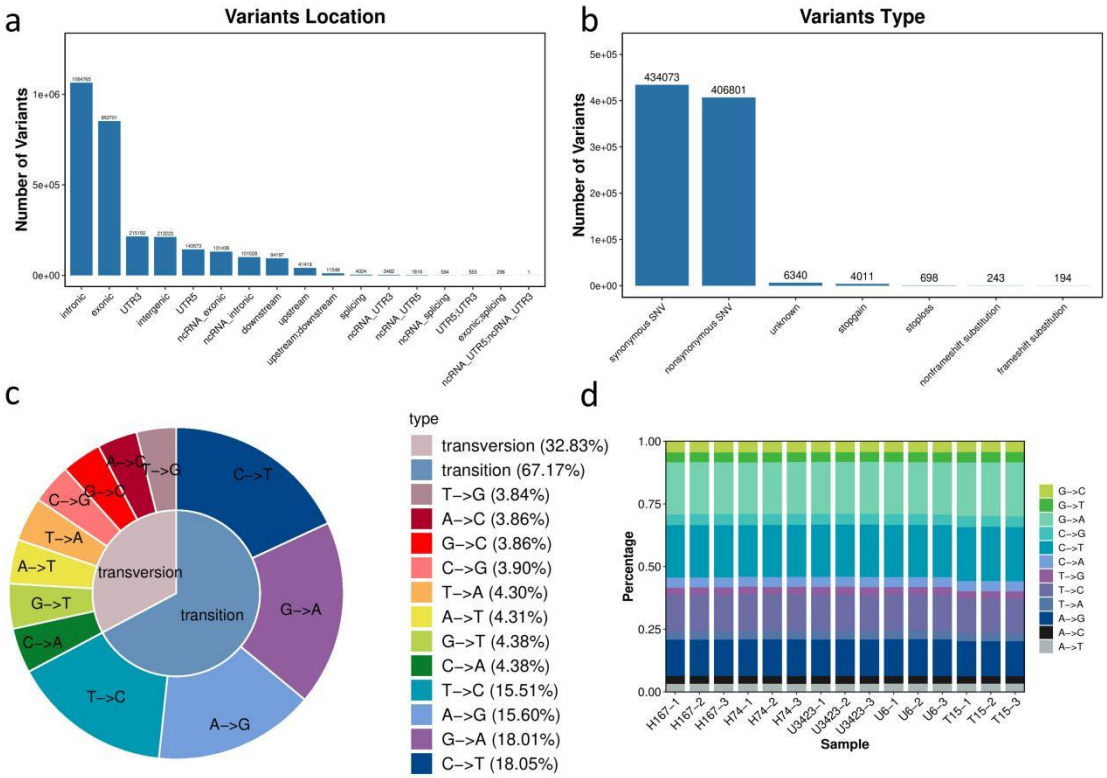


Figure S13 SNP/InDel analysis (Note: a: SNP location classification statistical chart; b: SNP functional classification statistical chart; c: SNP mutation statistical chart. Transversion refers to the substitution between purine and pyrimidine in base substitution (G<->T, C<->A, A<->T, C<->G), transition refers to the substitution of one purine with another purine or one pyrimidine with another pyrimidine (T<->C, A<->G). d: RNA Editing Classification Statistical Chart)

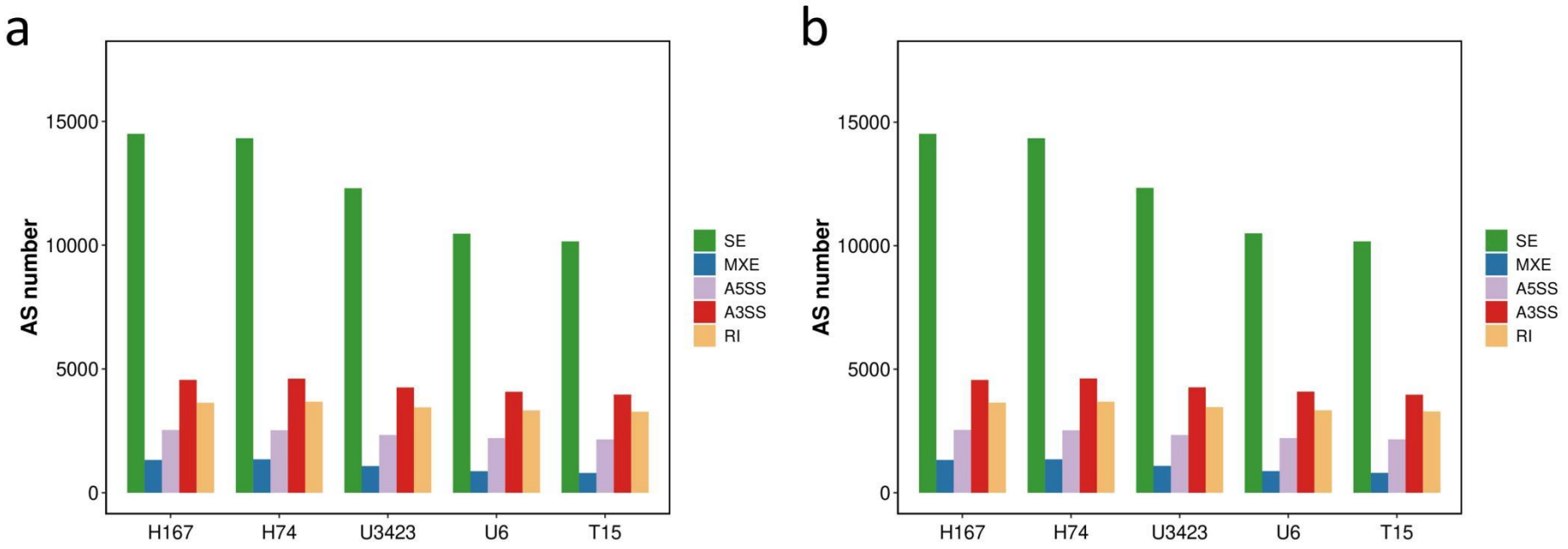
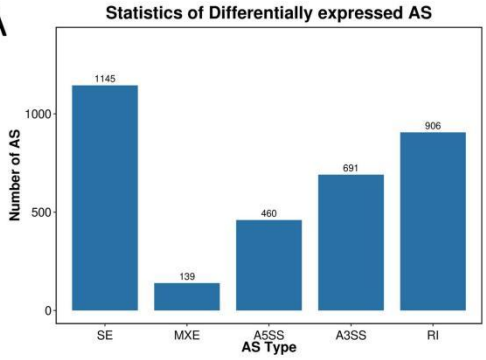


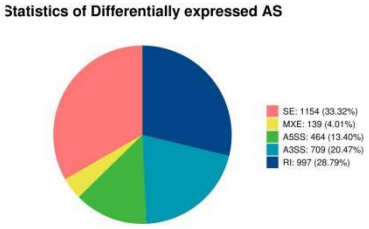
Figure S14 Variable cutting event statistics diagram (Note: The horizontal axis represents grouping, the vertical axis represents variable cutting quantity, and 5 columns correspond to 5 types of variable cutting events. a: JC Variable Shear Event Statistics Chart; b: JCEC Variable Shear Event Statistics Chart.)

A

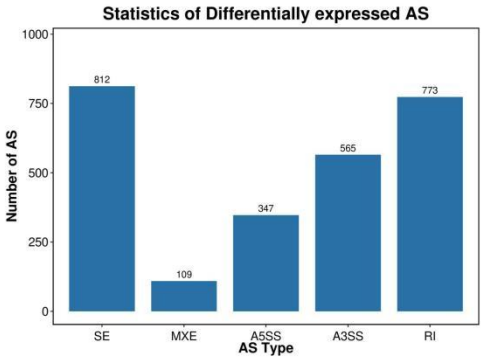
a

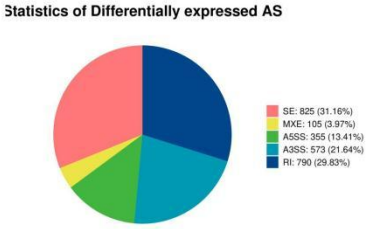

Statistics of Differentially expressed AS
Number of AS
1145
139
460
691
906
SE
MXE
A5SS
A3SS
RI
AS Type

B

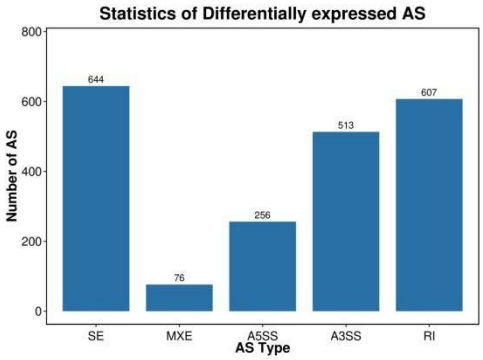

Statistics of Differentially expressed AS
SE: 1154 (33.32%)
MXE: 139 (4.01%)
A5SS: 464 (13.40%)
A3SS: 709 (20.47%)
RI: 997 (28.79%)

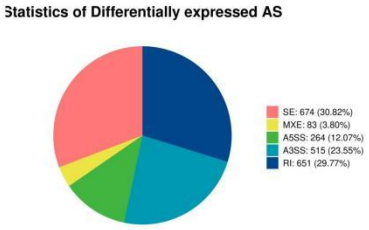
b


Statistics of Differentially expressed AS
Number of AS
812
109
347
565
773
SE
MXE
A5SS
A3SS
RI
AS Type

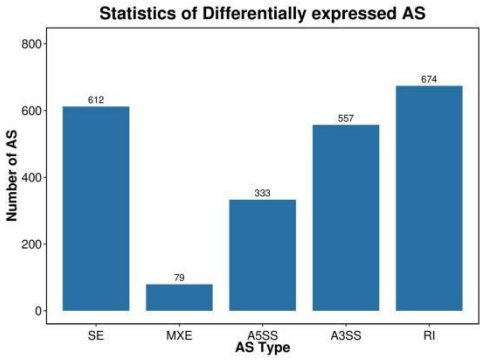

Statistics of Differentially expressed AS
SE: 825 (31.16%)
MXE: 105 (3.97%)
A5SS: 355 (13.41%)
A3SS: 573 (21.64%)
RI: 790 (29.83%)

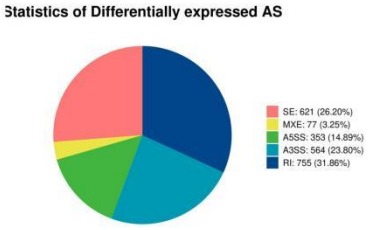
c


Statistics of Differentially expressed AS
Number of AS
644
76
256
513
607
SE
MXE
A5SS
A3SS
RI
AS Type

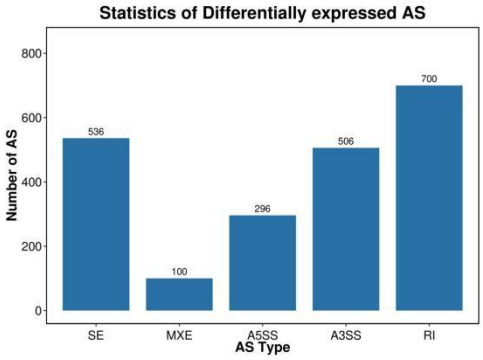

Statistics of Differentially expressed AS
SE: 674 (30.82%)
MXE: 83 (3.80%)
A5SS: 264 (12.07%)
A3SS: 515 (23.55%)
RI: 651 (29.77%)

d


Statistics of Differentially expressed AS
Number of AS
612
79
333
557
674
SE
MXE
A5SS
A3SS
RI
AS Type


Statistics of Differentially expressed AS
SE: 621 (26.20%)
MXE: 77 (3.25%)
A5SS: 353 (14.89%)
A3SS: 564 (23.80%)
RI: 755 (31.86%)

e


Statistics of Differentially expressed AS
Number of AS
536
100
296
506
700
SE
MXE
A5SS
A3SS
RI
AS Type

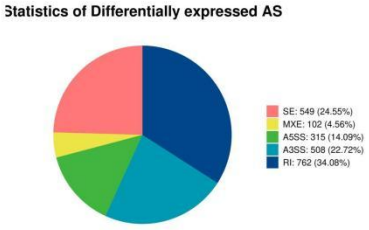

Statistics of Differentially expressed AS
SE: 549 (24.55%)
MXE: 102 (4.56%)
A5SS: 315 (14.09%)
A3SS: 508 (22.72%)
RI: 762 (34.08%)

Figure S15 Differential AS classification and quantity statistical diagram (Note: A: JC only difference AS classification and quantity statistical chart; The expression level of the alternative splicing event Exon Skipping Isoform in the sample (the result of aligning reads between two constitutive exons, i.e. reads spanning upstream and downstream constitutive exons). B: JC+reads OnTarget difference AS classification and quantity statistics chart. The expression level of the alternative splicing event Exon Inclusion Isoform in the sample (reads the alternative splice junction between upstream constitutive exons and skipping exons; The reads are completely aligned to the skip type exon; Reads spans across the skip type exon and downstream constitutive exon alternative splice junction.) The number of reads for AS events in the sample is calculated using reads on target and junction counts. a: H74 vs U6; b: H74 vs U3423; c: H167 vs H74; d: H167 vs T15; e: H167 vs U3423.)

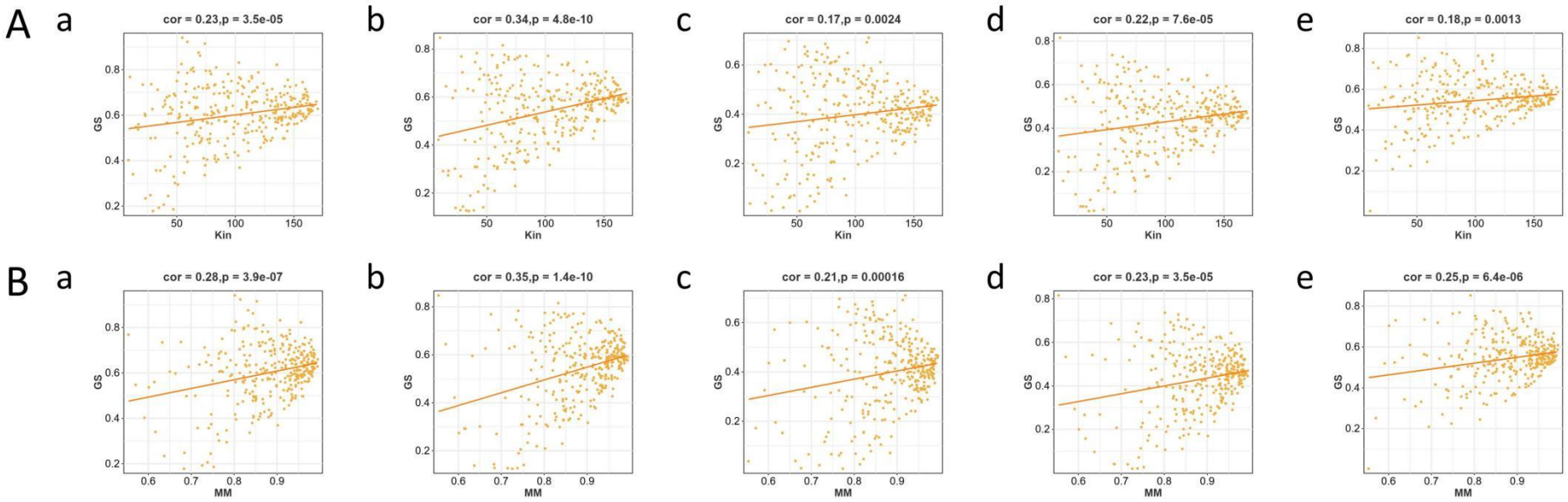


Figure S16 Scatter plot of correlation with GS (Note: A/B: Scatter plot of K.within/MM and GS correlation for the most positively correlated (left) and most negatively correlated (right) modules. The horizontal axis represents the K.within/MM value, the vertical axis represents the GS value, the dots represent genes, and the diagonal line represents the fitting line of correlation. a: HT; b: DBH; c: SUR; d: VOL; e: SS.)

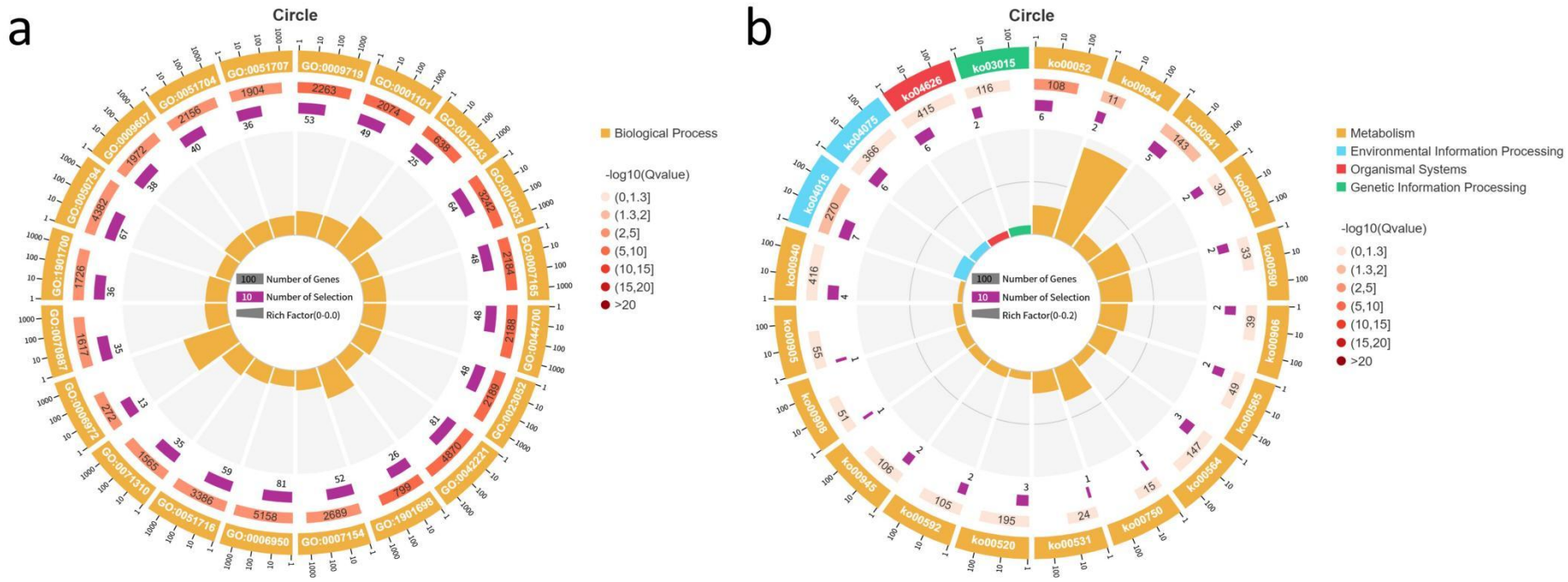


Figure S17 Enrichment circle diagram - WGCNA (Note: Tan module; a: GO; b: KEGG)

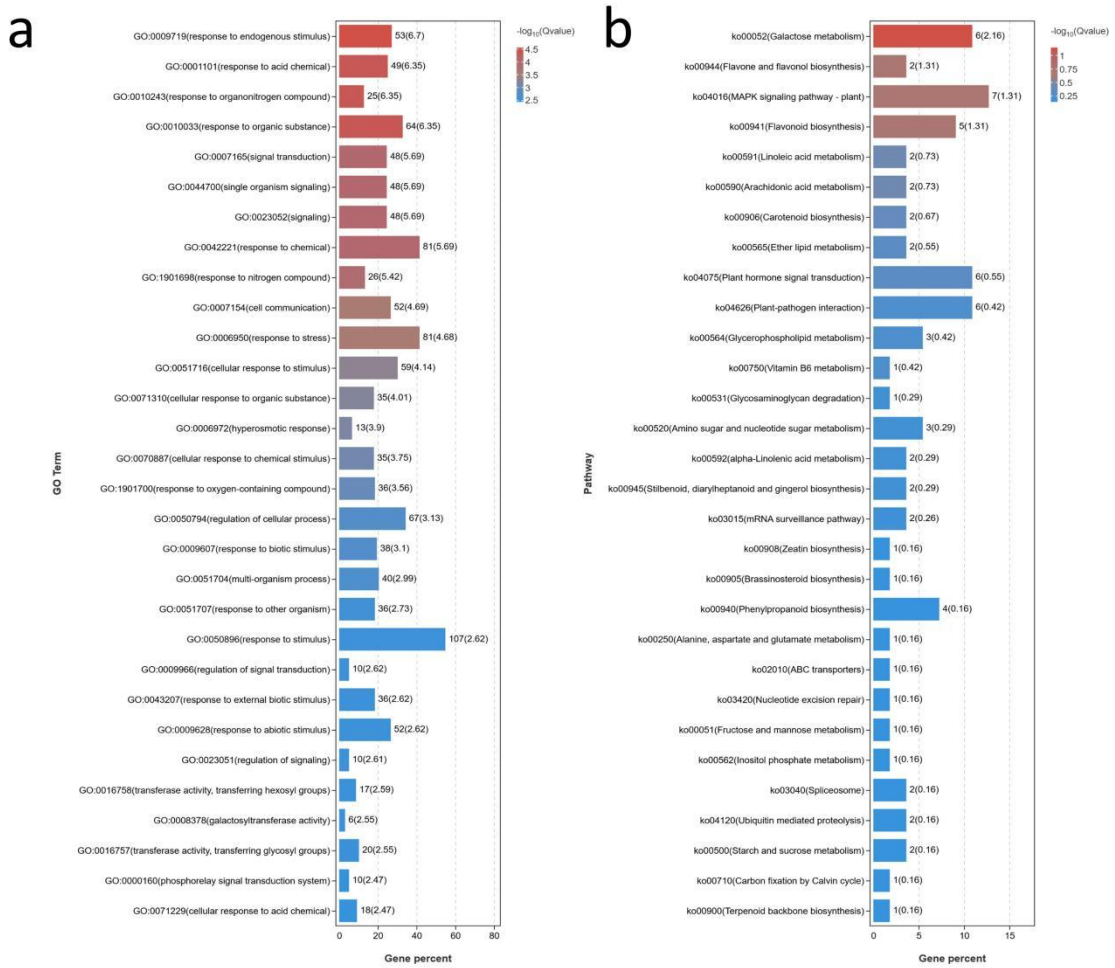


Figure S18 Enrichment bar diagram - WGCNA (Note: Tan module; a: GO; b: KEGG)

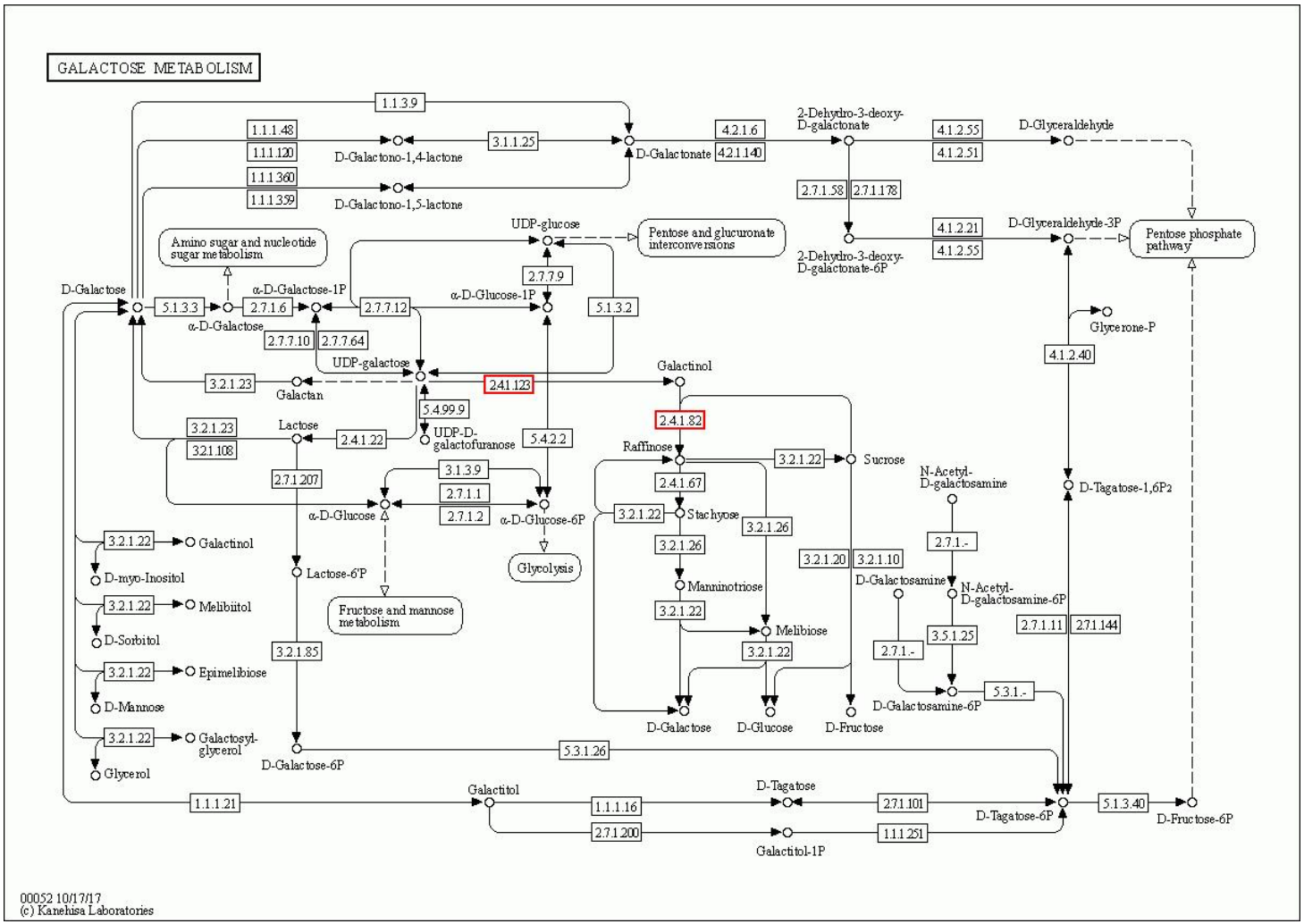


Figure S19 Detailed information on Pathways in KEGG database - WGCNA (Note: Tan module Galactose metabolism(KO00052))

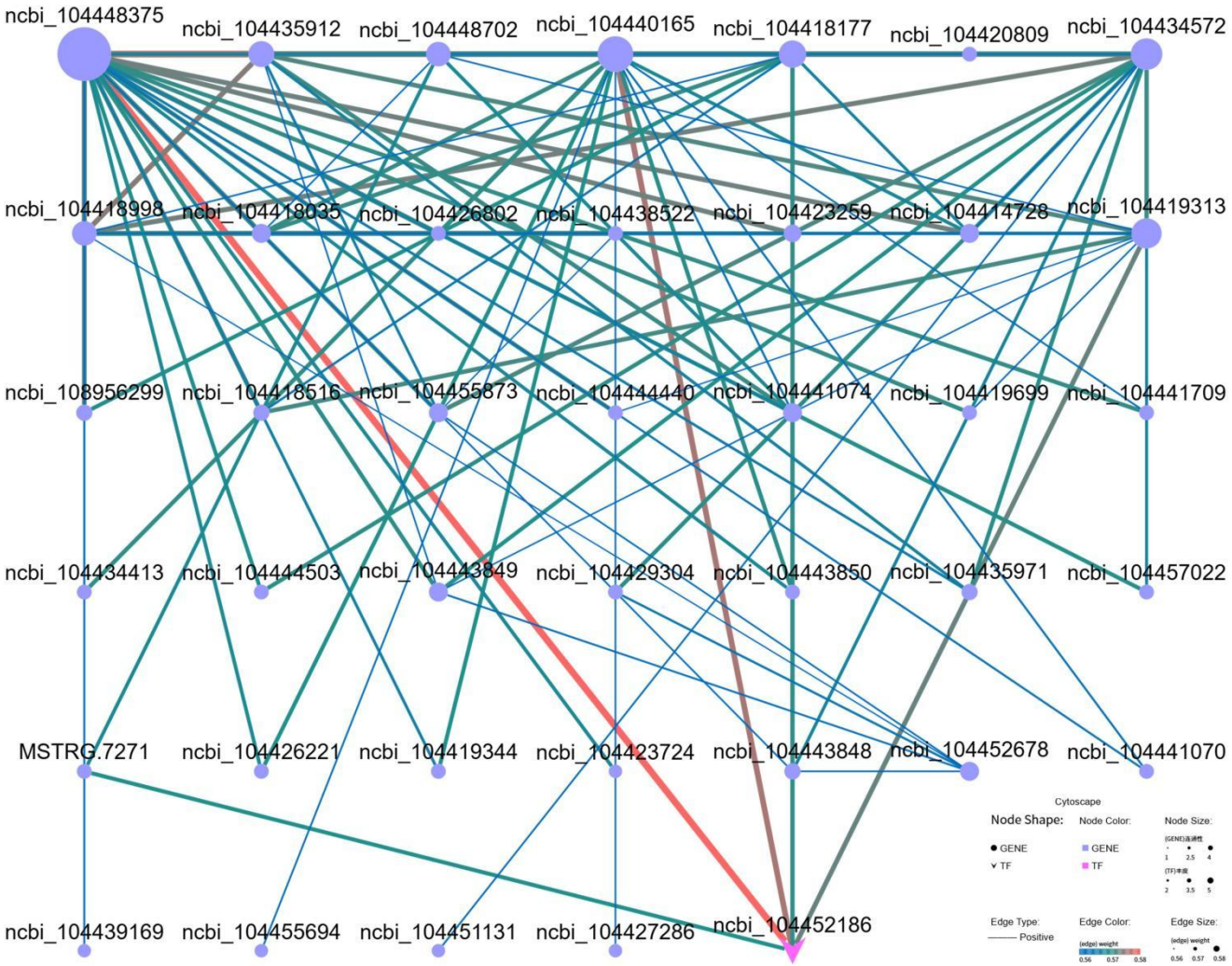


Figure S20 Gene regulatory network diagram (Note: Tan module; Top 100 Connectivity Relationships)

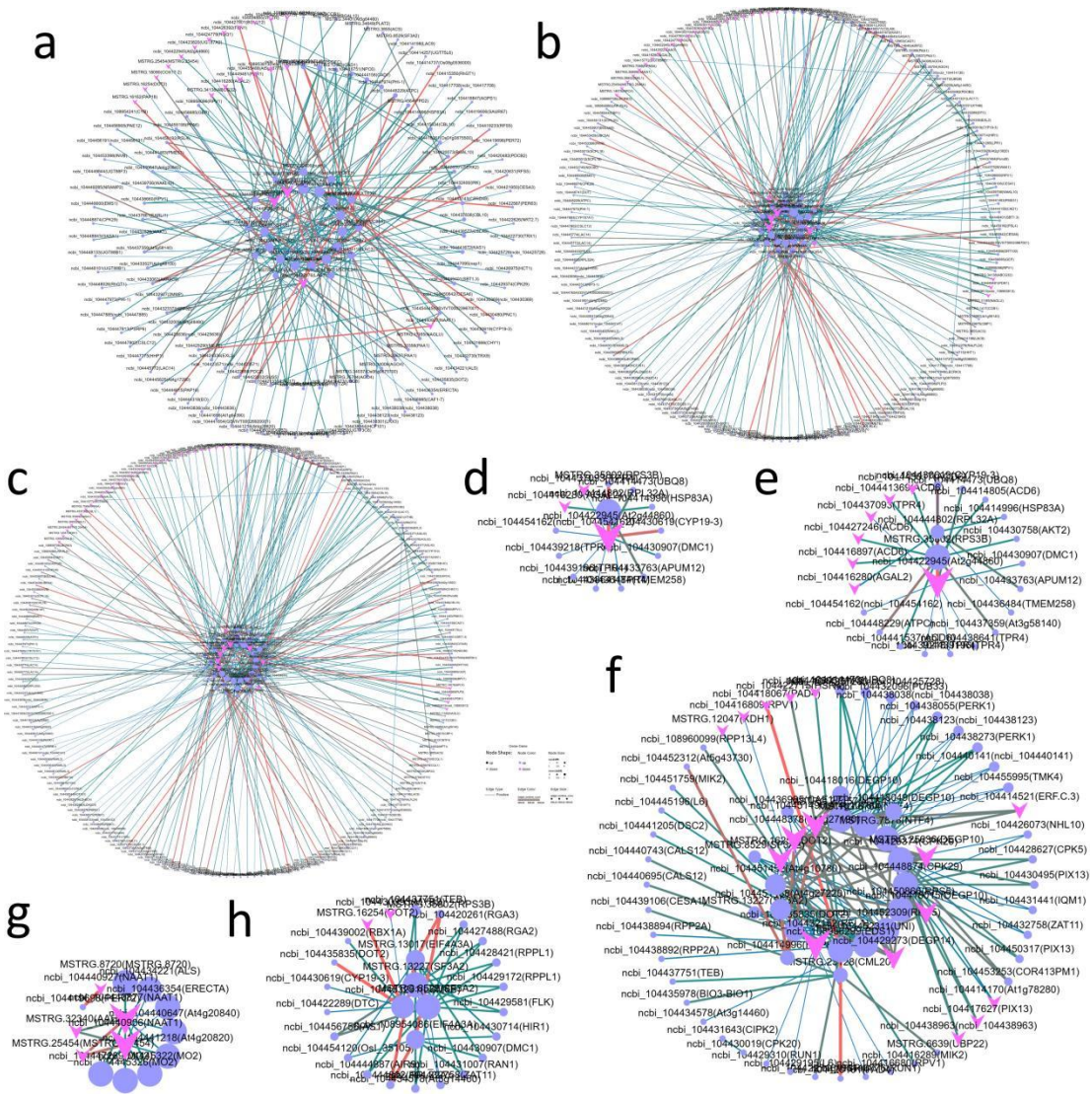


Figure S21 Preliminary target gene PPI network (Note: concentric diagram. a: extracellular region (GO:0005576); b: external encapsulating structure (GO:0030312); c: cell periphery (GO:0071944); d: large ribosomal subunit (GO:0015934); e: ribosomal subunit (GO:0044391); f: Plant-pathogen interaction (KO04626); g: Phenylalanine, tyrosine and tryptophan biosynthesis (KO00400); h: Spliceosome (KO03040).)

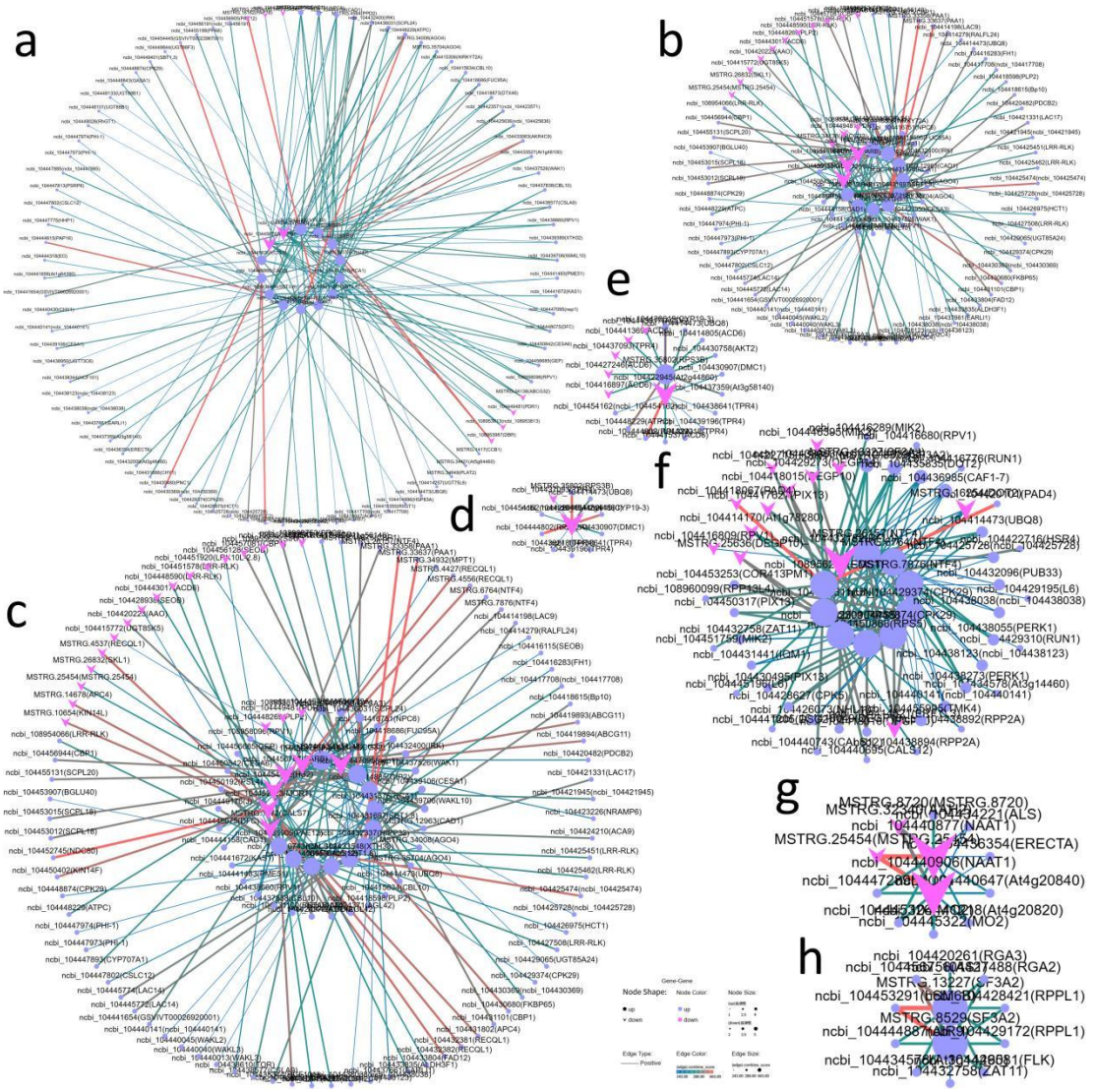


Figure S22 Target gene PPI network (Note: concentric diagram. a: extracellular region (GO:0005576); b: external encapsulating structure (GO:0030312); c: cell periphery (GO:0071944); d: large ribosomal subunit (GO:0015934); e: ribosomal subunit (GO:0044391); f: Plant-pathogen interaction (KO04626); g: Phenylalanine, tyrosine and tryptophan biosynthesis (KO00400); h: Spliceosome (KO03040).)