CLAIMS

1. A method for producing a nucleic acid library that is comprised of a plurality of expression control sequences that are configured to modulate the expression of coding sequence operably linked thereto in a plant cell, the method comprising:

performing an *in silico* analysis of a genome for a plant subject, in which the *in silico* analysis includes identification of a plurality of sequences that collectively define a promoterome of the plant subject, in which each sequence within the promoterome is comprised of an expression control region that extends from the start codon of an open reading frame to around 100 kilobases 5' and/or 3' to the start codon;

obtaining a transcriptome in the form of mRNA expression data for the plant subject;

undertaking an analysis using a sequence-based modelling algorithm in order to provide a prediction value of the level of expression of each protein coding gene comprised within the transcriptome and linking the value to a sequence for a corresponding expression control region comprised within the promoterome;

generating a plurality of non-wild type sequence designs for expression control sequences that are most likely to provide a desired expression profile for an operably linked coding sequence, in which the plurality of non-wild type sequence designs are informed by the prediction value;

synthesising a plurality of non-wild type expression control sequences that correspond to the plurality of non-wild type sequence designs; and

generating a nucleic acid sequence library that comprises the plurality of non-wild type expression control sequences.

2. A method according to Claim 1, in which the expression control region comprises one or more sequence selected from the group consisting of: a promoter; an enhancer; a silencer; a transcription factor binding site; an intron; a transgenic sequence; and a transposon.

3. A method according to Claim 1 or 2, in which the promoterome comprises all or a part of a protein coding region.

4. A method according to Claim 1 or 2, in which the promoterome does not comprise a protein coding region.

5. A method according to any preceding claim, in which each sequence within the promoterome is comprised of an expression control region that extends from the start codon of an open reading frame to less than 100 kilobases 5' and/or 3' to the start codon, suitably less than 50 kilobases 5' and/or 3' to the start codon.

6. A method according to any preceding claim, in which the analysis using a sequence-based modelling algorithm comprises training an artificial intelligence (AI) or machine learning (ML) model.

7. A method according to Claim 6, in which the AI or ML model comprises a sequence prediction algorithm.

8. A method according to Claim 7, sequence prediction algorithm is selected from: an artificial neural network (ANN) algorithm such as those selected from the group consisting of: a convolutional neural network (CNN); and a recurrent neural network (RNN), including a bidirectional RNN, masked language model and a transformer network.

9. A method according to any preceding claim, in which the desired expression profile

prioritises expression control sequences that provide a higher level of expression than the prediction value; or

provide a lower level of expression than the prediction value; or

in which the plurality of non-wild type sequence designs are constrained by a minimal change requirement

10. A method according to any preceding claim, in which the method further comprises validating one or more of the plurality of non-wild type expression control sequences *in vivo*.

11. A method according to Claim 10, in which the *in vivo* validation comprises cloning one or more of the plurality of non-wild type expression control sequences into one or more plasmids that comprise a reporter gene whose expression is operably linked to the corresponding non-wild type expression control sequence.

12. A method according to Claim 11, in which the *in vivo* validation comprises undertaking a massively parallel reporter assay (MPRA).

13. A method according to Claim 12, in which the MPRA comprises a fluorescence activated cell sorting (FACS) analysis.

14. A method according to Claim 11, in which each of the one or more plasmids that comprise a reporter gene further comprises a unique molecular identifier (UMI) sequence.

15. A method according to Claim 14, in which the *in vivo* validation comprises undertaking a transcriptional activity assay that measures the amount of the reporter gene mRNA produced and correlates the amount to the corresponding non-wild type expression control sequence via the UMI.

16. A method according to any of Claims 10 to 15, in which the *in vivo* validation is carried out within a plant protoplast within a plant cell within a plant; or within a part of a plant.

17. A method according to any preceding claim, in which the plant subject comprises a plant species or plant variety.

18. A method according to Claim 17, in which the plant species or plant variety is selected from the group consisting of: *Solanum spp. (e.g. S. lycopersicum, S. tuberosum, S. melongena, S. muricatum, S. betaceum); Brassica spp. (e.g. B.oleracea, B.napobrassica, B.napus, B.cretica, B.rupestris and B.rapa); Capsicum spp. (e.g. C. annuum, C. baccatum, C. chinense, C. frutescens, C. pubescens); Lupinus spp. (e.g. L. angustifolius, L. albus, L. mutabilis and L. luteus); Phaseolus spp. (e.g. P. acutifolius, P. coccineus, P. lunatus, P. vulgaris and P. dumosus); Vigna spp (e.g. V. aconitifolia, V. angularis, V. mungo, V. radiata, V. subterranea and V. unguiculata); Vicia faba; Cicer arietinum, Pisum sativum, Lathyrus spp. (e.g. L. sativus and L. tuberosus); Lens spp. (e.g. L. culinaris and L. esculenta); Glycine max; Psophocarpus; Cajanus cajan; Arachis hypogaea; Lactuca spp. (e.g. L. sativa, L. serriola, L. saligna, L. virosa and L. taterica); Asparagus officinalis; Apium graveolens; Allium spp. (e.g. A. cepa, A. oschaninii, A. ampeloprasum, A. wakegi, A. porrum, A. sativum and A. schoenoprasum); Beta vulgaris; Cichorium intybus; Taraxacum officinale, Eruca spp. (e.g. E. vesicaria and E. sativa); Cucurbita spp. (e.g. C. argyosperma, C. digitata, C. pepo, C. moschata, C. ecuadorensis, C. ficifolia, C. foetidissima, C. galeottii, C. lundelliana, C. maxima, C. moshata, C. pedatifolia, C. radicans); Spinacia oleracea; Nasturtium officinale; Cucumis spp. (e.g. C. sativus, C. melo, C. hystrix, C. picrocarpus and C. anguria); Olea europaea; Daucus carota; Ipomoea batatas; Ipomoea eriocarpa; Manihot esculenta; Zingiber officinale; Armoracia rusticana; Helianthus spp. (e.g. H. annuus and H. tuberosus); Cannabis*

spp. (e.g. C. sativa and C. indica); Pastinaca sativa; Raphanus sativus; Curcuma longa; Dioscorea spp. (e.g. D. rotundata, D. alata, D. polystachya, D. bulbifera, D. esculenta, D. dumetorum, D. trifida and D. cayennensis); Piper spp. (e.g. P. aduncum, P. guineense and P. nigrum); Zea spp. (e.g. Z. mays and Z. diploperennis); Hordeum spp. (e.g. H. vulgare, H. pusillum, H. murinum, H. marinum, H. jubatum and H. intercedens); Gossypium spp. (e.g. G. hirsutum, G. barbadense, G. arboreum and G. herbaceum); Triticum spp. (e.g. T. aestivum and T. timopheevii); Vitis vinifera; Prunus sp. (e.g. P. avium, P. armeniaca, P. cerasifera, P. cerasus, P. domestica, P. persica and P. dulcis); Malus domestica; Pyrus spp. (e.g. P. communis, P. cordata and P. pyrifolia); Fragaria vesca and Fragaria × ananassa; Rubus idaeus; Saccharum officinarum; Sorghum saccharatum; Musa balbisiana and Musa x paradisiaca; Oryza sativa; Nicotiana tabacum; Arabidopsis thaliana; Citrus spp. (e.g. C. × aurantiifolia, C. × aurantium, C. × latifolia, C. × limon, C. × limonia, C. × paradisi, C. × sinensis and C. × tangerina); Populus spp. (e.g. P. tremula, P. balsamifera and P. tomentosa); Tulipa gesneriana; Medicago sativa; Abies balsamea; Avena orientalis; Bromus mango; Calendula officinalis; Chrysanthemum balsamita; Dianthus caryophyllus; Eucalyptus spp. (e.g. E. leucoxylon, E. maculata, E. polybractea, E. sargentii); Impatiens biflora; Linum usitatissimum; Lycopersicon esculentum; Mangifera indica; Nelumbo spp. (e.g. N. nucifera and N. pentapatala); Poaceae spp.; Secale cereale; Tagetes erecta; and Tagetes minuta.

19. A nucleic acid library that comprises a plurality of non-wild type expression control sequences that are configured to modulate the expression of coding sequence operably linked thereto within a plant cell, in which the plurality of non-wild type expression control sequences are generated by the method as defined in any of Claims 1 to 18.