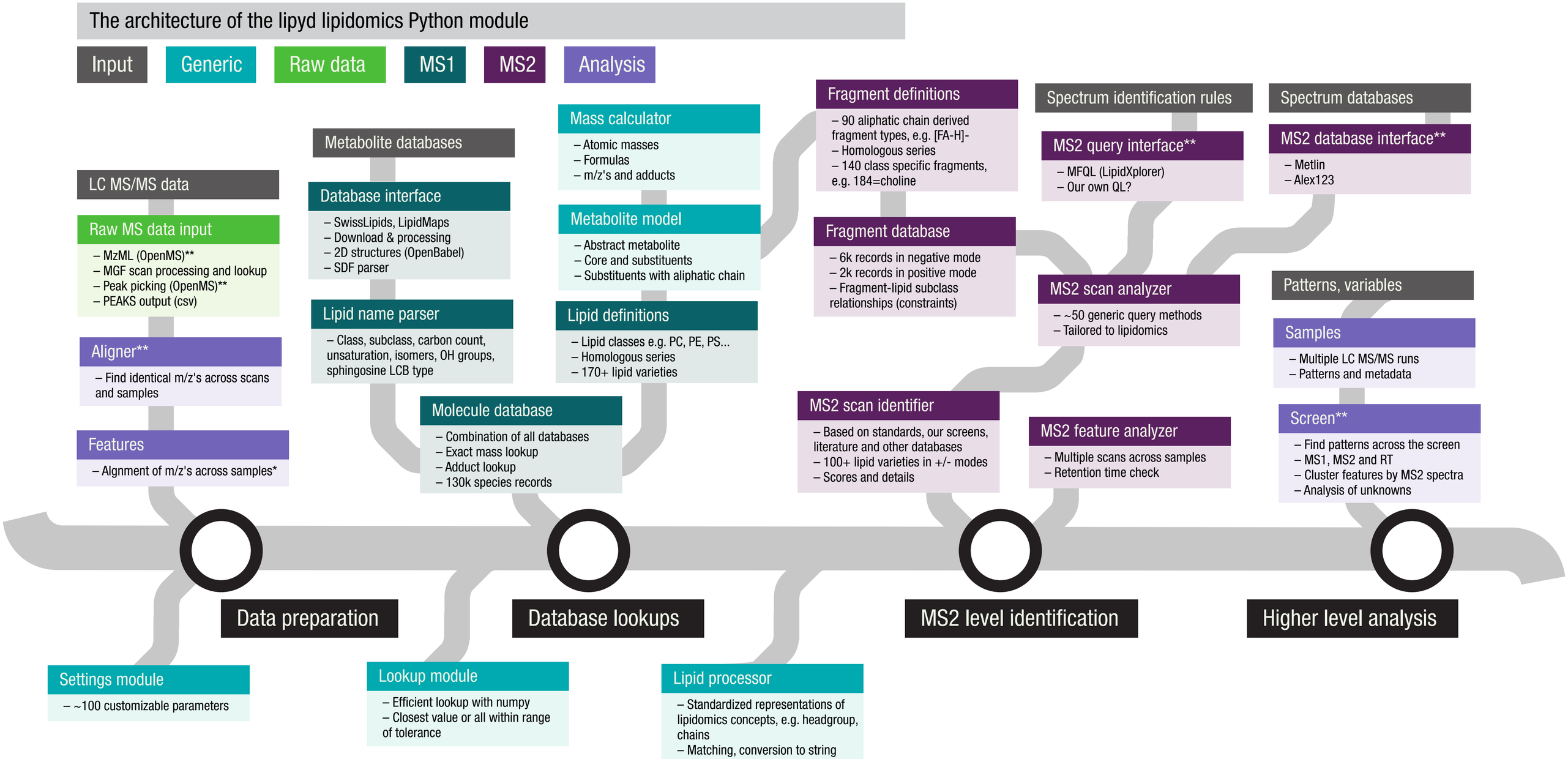
The architecture of the lipyd lipidomics Python module
Input
Generic
Raw data
MS1
MS2
Analysis
LC MS/MS data
Raw MS data input
– MzML (OpenMS)**
– MGF scan processing and lookup
– Peak picking (OpenMS)**
– PEAKS output (csv)
Aligner**
– Find identical m/z's across scans and samples
Features
– Alignment of m/z's across samples*
Metabolite databases
Database interface
– SwissLipids, LipidMaps
– Download & processing
– 2D structures (OpenBabel)
– SDF parser
Lipid name parser
– Class, subclass, carbon count, unsaturation, isomers, OH groups, sphingosine LCB type
Mass calculator
– Atomic masses
– Formulas
– m/z's and adducts
Metabolite model
– Abstract metabolite
– Core and substituents
– Substituents with aliphatic chain
Lipid definitions
– Lipid classes e.g. PC, PE, PS...
– Homologous series
– 170+ lipid varieties
Molecule database
– Combination of all databases
– Exact mass lookup
– Adduct lookup
– 130k species records
Fragment definitions
– 90 aliphatic chain derived fragment types, e.g. [FA-H]-
– Homologous series
– 140 class specific fragments, e.g. 184=choline
Fragment database
– 6k records in negative mode
– 2k records in positive mode
– Fragment-lipid subclass relationships (constraints)
Spectrum identification rules
MS2 query interface**
– MFQL (LipidXplorer)
– Our own QL?
Spectrum databases
MS2 database interface**
– Metlin
– Alex123
MS2 scan analyzer
– ~50 generic query methods
– Tailored to lipidomics
MS2 scan identifier
– Based on standards, our screens, literature and other databases
– 100+ lipid varieties in +/- modes
– Scores and details
MS2 feature analyzer
– Multiple scans across samples
– Retention time check
Patterns, variables
Samples
– Multiple LC MS/MS runs
– Patterns and metadata
Screen**
– Find patterns across the screen
– MS1, MS2 and RT
– Cluster features by MS2 spectra
– Analysis of unknowns
Data preparation
Database lookups
MS2 level identification
Higher level analysis
Settings module
– ~100 customizable parameters
Lookup module
– Efficient lookup with numpy
– Closest value or all within range of tolerance
Lipid processor
– Standardized representations of lipidomics concepts, e.g. headgroup, chains
– Matching, conversion to string